Spatial structure of city population growth

Sandro M. Reia, P. Suresh C. Rao, Marc Barthelemy, Satish V. Ukkusuri

nature communications

Article

https://doi.org/10.1038/s41467-022-33527-y

Spatial structure of city population growth

Received: 29 March 2022

Accepted: 22 September 2022

Published online: 08 October 2022

Check for updates

Sandro M. Reia [1], P. Suresh C. Rao [1,2], Marc Barthelemy [3,4] & Satish V. Ukkusuri [1]

We show here that population growth, resolved at the county level, is spatially heterogeneous both among and within the U.S. metropolitan statistical areas. Our analysis of data for over 3,100 U.S. counties reveals that annual population flows, resulting from domestic migration during the 2015–2019 period, are much larger than natural demographic growth, and are primarily responsible for this heterogeneous growth. More precisely, we show that intra-city flows are generally along a negative population density gradient, while inter-city flows are concentrated in high-density core areas. Intra-city flows are anisotropic and generally directed towards external counties of cities, driving asymmetrical urban sprawl. Such domestic migration dynamics are also responsible for tempering local population shocks by redistributing inflows within a given city. This spill-over effect leads to a smoother population dynamics at the county level, in contrast to that observed at the city level. Understanding the spatial structure of domestic migration flows is a key ingredient for analyzing their drivers and consequences, thus representing a crucial knowledge for urban policy makers and planners.

Research on city population growth has a long history with statistical regularities among the cities, as identified in early seminal works by Auerbach[1] and later by Zipf[2]. Random demographic growth was generally considered[3,4] as the main source of the population growth dynamics of cities. However, recently[5,6] city population growth was shown to result from a combination of random demographic growth and, more importantly, inter-city flows from domestic migration that are broadly distributed according to a power law[5]. These flows are triggered by socioeconomic changes and can dramatically alter the trajectory of the population growth of a city[7,8].

Inter-city flows from domestic migration play a crucial role in the evolution of the system of cities at a country scale and its analysis is fundamental to understand the temporal and spatial evolution of cities. More specifically, the structure of household domestic migration provides insights on regions that are more likely to grow, which is usually accompanied by various externalities, such as traffic congestion[9–11], air pollution[12,13] and socioeconomic inequality[14]. Extreme flows lead to unprecedented population growth that is usually more expansive than compact[15–17], thus understanding the structure of domestic migration flows at intra- and inter-city scales help to plan for various unforeseen problems and to devise mitigation strategies. This is particularly important and well known for the suburban and fringe area urbanization[18], which have their own planning challenges and peculiarities[19].

The dynamics of city population growth is usually studied at the city level, neglecting the intra-city spatial structure of migratory flows. Spatial heterogeneity of cities is well known, evident in consistent patterns, among others, nonlinear decrease in population density with increasing distance from the dense urban core[20], fractal urban morphology[21], spatial structure of urban heat islets[22]. Other urban studies focused on emergence of inequalities among neighborhoods and infrastructure development[8,23], and on topological properties of flow networks[24,25].

In this paper, we study the spatial heterogeneity in city population growth by conducting an analysis of the most recent American Community Survey (ACS) 5-year county-to-county migration flow files. We focus on the origin and destination counties of the domestic migration flows, revealing spatial variations of components of population growth at the county level within metropolitan statistical areas. For this reason, we interchangeably use the terms city and metropolitan statistical

[1]Lyles School of Civil Engineering, Purdue University, 550 Stadium Mall Avenue, West Lafayette 47907 IN, USA. [2]Agronomy Department, Purdue University, 915 W State Street, West Lafayette 47907 IN, USA. [3]Université Paris-Saclay, CNRS, CEA, Institut de physique théorique, Gif-surYvette 91191, France. [4]Centre d'Analyse et de Mathématique Sociales, CNRS/EHESS, Paris 75006, France. e-mail: sukkusur@purdue.edu

area (MSA). Our goal here is to examine generalized patterns across cities, despite their specific differences. We show that inter-city flows are more likely to occur between core counties (the core county has the highest population density in a city), and intra-city flows are more likely to follow an outward radial direction (i.e., there is a trend towards exterior and lower density counties). Moreover, flows to/from micropolitan statistical areas and rural counties are more likely to happen at the external regions of cities, and international inflows are more likely to be found at core counties of large cities (see Fig. 1).

Results

Overview of U.S. domestic migration flows

The most recent ACS county-to-county flow dataset[26] reports that about 45.6 million people migrated to the U.S. per year during the period 2015–2019, which corresponds to 14.2% of the U.S. population[27]. Approximately 43.5 million annual moves corresponded to domestic migration (moves within the U.S.[28]), while 2.1 million accounted for inflows of individuals from other countries (viz. international immigration).

With respect to domestic migration, 25.7 million people per year migrated within the same county, thus showing that the highest share of domestic flows (59%) is intra-county. Annually, about 10.4 people moved between different counties within the same state, thus intra-state flows account for 24% of the domestic migration (Supplementary Fig. 1), mainly driven by the search for more affordable housing, better jobs, and for family reasons such as change in marital status[29]. Long distance moves, captured by inter-state flows, represent the remaining 17% of domestic flows, which comprises about 7.5 million moves per year. Here, we will refer to these domestic migration flows as inflows or outflows, and netflows (inflows-outflows).

The United States Office of Management and Budget (OMB) classifies counties as metropolitan, micropolitan, or neither[30]. A metropolitan statistical area contains a core urban area of at least 50,000 population. A metro area represents a functional delineation of an urban area with a network of strong socioeconomic ties, and provision of infrastructure services[31–33]. A micropolitan statistical area contains an urban core of at lest 10,000 but less than 50,000 inhabitants. There are over 380 metropolitan statistical areas in the U.S., each composed of one or more counties, accounting for about 86% of the total U.S. population and comprising approximately 28% of the land area of the country. For this reason, our analysis focuses on the growth dynamics of MSA counties. Supplementary Fig. 2 shows the 3141 counties (administrative subdivisions of the states) in the U.S., comprising about 321 million inhabitants in the starting year of the ACS 5-Year survey period (2015–2019) of our analysis[26].

Population growth has two components, namely natural growth and migration. Natural growth accounts for births minus deaths, and migration comprises domestic and international migration. With recent trends showing that births and natural increase have declined in the U.S. and in recent years contribute less to overall city population growth[34,35], migration patterns become more relevant to the study of city population growth. Because the ACS flow files contain international inflows only, the relative importance of migrations on population growth is here addressed by $x = |\text{Inflows} - \text{Outflows}|/|\text{Births} - \text{Deaths}|$ (Supplementary Figs. 3, 4), which is the ratio between domestic netflows and natural growth. The statistical distribution of this quantity computed for all U.S. counties is well fitted by a log-normal distribution, and shows that $x \geq 1$ for 76.5% of counties. For most counties, domestic migration dominates population growth, and understanding the spatial structure of domestic netflows (and their distribution within a city) is crucial to the comprehension of the mechanisms behind the heterogeneity of city population growth.

At this spatial granularity, we observe a strong heterogeneity among the U.S. counties (Supplementary Fig. 2) for the period 2015 – 2019, along with examples of specific MSAs. In particular, the relative dispersion of counties relative growth due to netflows is higher than one for about 85% of the metro areas, indicating a large heterogeneity within the same city and pointing towards the spatial structure of domestic migration. The observed difference in the netflows stresses the relevance of our approach: counties belonging to the same city

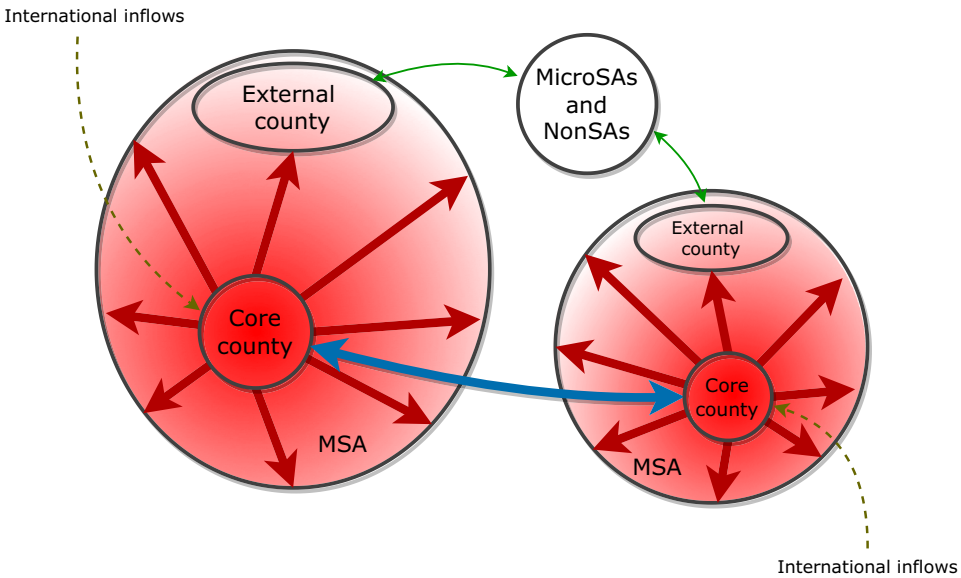

Fig. 1 | Schematic representation of the dominant migratory trends that contribute to the heterogeneous population growth of cities. Core counties are more likely to receive inflows from core counties of other cities than from external counties (blue arrows). Flows to and from micro and non-statistical areas are more likely to be found at the external counties of a city (green arrows). Intra-city flows (red arrows) indicate vectors of redistribution of people within the city, and have an outwards radial direction: people move from central counties with larger population density to external counties with lower densities. International inflows (dashed arrows), which scale superlinearly with city population, are more likely to be directed to the core counties of large cities. The resulting spatial heterogeneity is depicted by the background color, in which the red intensity is proportional to the population density. The width of the arrows is proportional to the intensity of the flows.

may have specific growth rates due to population flow patterns, thus indicating preferential flow destinations and pinpointing the direction in which the city has expanded.

Heterogeneity of inter- and intra-city flows

Inter-city flows represent the major component of the total flows (~55%), while intra-city flows represent ~25%. Flows between metro and micro areas, and between metro and non-statistical areas are the smallest components, with ~13% and ~7%, respectively. Given that about 80% of the domestic migration are composed of intra- and inter-city flows, we will focus our attention on describing the structure of intra- and inter-city flows, but in the Supplementary Information we offer a brief analysis of flows between metro and micro areas, and between metro and non-statistical areas.

Inter-city flows are not uniform across the U.S. cities. The most intense annual netflows (>2000 people per year), accounting for approximately 17% of the entire inter-city U.S. netflows, are mainly from New York and Chicago to California and Florida (Fig. 2), and from Los Angeles to neighboring cities. Notably, netflows among the Midwestern cities are mostly negative and below the threshold we set. These flows are mainly responsible for increasing or decreasing the population of a given city. Intra-city flow patterns, illustrated with the 7 most populous U.S. cities with more than 5 counties, are also non-uniform.

Our analysis reveals that city centers (defined as the core county with the highest population density) are more likely to have negative netflows, indicating that people are leaving the central regions of cities. The arrows in Fig. 2 indicate the direction of the most intense netflows, supporting this finding and highlighting that there is a trend of people moving from internal to external regions, contributing to population growth and spatial expansion of U.S. cities. In fact, we found no correlation between relative population growth (viz. population growth by county size) and distance from the core county (Supplementary Fig. 5A) for the 46 cities with more than 5 counties, with relative growth about 0.03 ± 0.05. On the other hand, we found that relative natural growth (Supplementary Fig. 5B) is negatively correlated with the distance to core county, thus natural growth is less relevant as a component of growth in the outer regions of cities. Consequently, our results show that not only the contribution of each component of growth changes with distance to core county, but also that the internal redistribution of people is an important mechanism of growth, mainly in the external counties.

We also examined variability in inter- and intra-city flows within the 50 states (Supplementary Fig. 6). Total flows within a state increase, as expected, with the state population. Two special cases are, however, of interest: (1) two states (Vermont and Rhode Island) with small populations have only one MSA, in which case within-state inter-city flows are zero; and (2) nearly 40%, or 149, of MSAs have only one county, in which case intra-city flows could not be estimated. For all other cases, we observe on average an equal split between inter- and intra-city flows, but with considerable variability among the states, with a mean about 0.5 and standard deviation about 0.2. A generalization of the intra- and inter-city migratory patterns for all 46 cities with more than 5 counties shows that the percentage of migrants from intra- and inter-city flows are of the same order of magnitude (Fig. 3).

Apart from the core county, flows from the same city correspond to about 50% of the inflow of people in the counties, presenting a slightly positive correlation with their distance from the city center (Fig. 3A). The low percentage for the core county indicates that it is not the major destination of flows from the same city. The percentage of

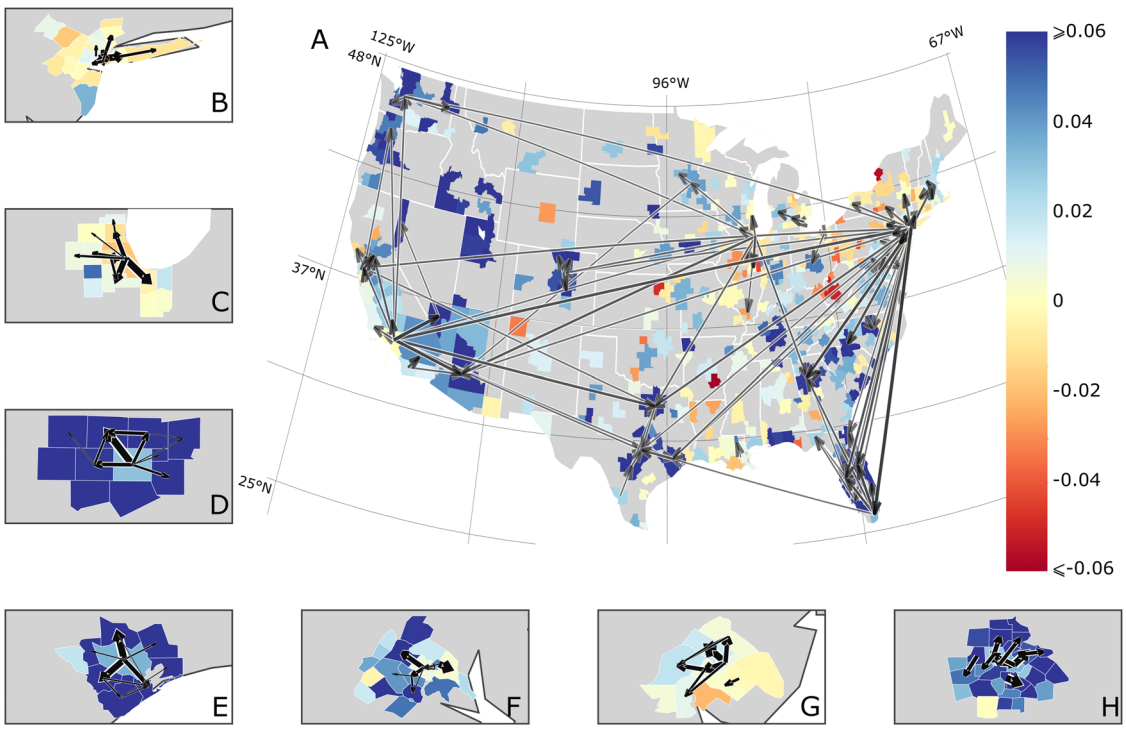

Fig. 2 | Heterogeneity of inter- and intra-city netflows. The map (**A**) suggests that the domestic redistribution of people between different U.S. metro areas are non-uniform: the black arrows, indicating the direction of the most intense inter-city netflows (higher than 2000 people per year), reveal migration trends from northern and eastern cities to western and southern regions. Cities (composed of one or more counties) are colored according to the relative growth (viz. population growth adjusted by population) of the whole MSA during the 2015–2019 period, and the black intensity and the thickness of the arrows are proportional to the netflows. Alaska and Hawaii are not shown. Panels (**B**–**H**), which are close-up of New York (**B**), Chicago (**C**), Dallas (**D**), Houston (**E**), Washington D.C. (**F**), Philadelphia (**G**), Atlanta (**H**), suggest that the most intense intra-city netflows are oriented radially outwards: people are moving from core to external counties. Here, counties are colored according to their relative growth in the 2015–2019 period and the width of the arrows is proportional to the netflows between origin and destination counties.

https://doi.org/10.1038/s41467-022-33527-y

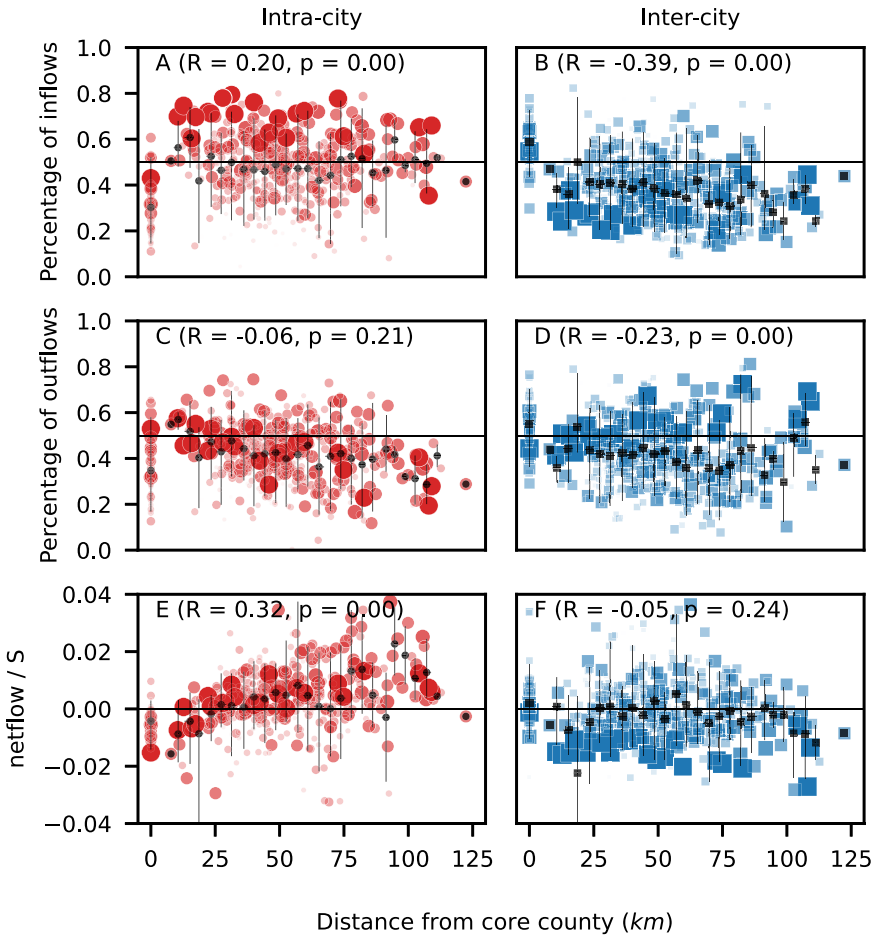

Fig. 3 | Roles of intra- and inter-city flows in driving the heterogeneous population growth of cities. We define the core county as the one with the highest population density, and we plot the percentage of inflows due to intra- (**A**) and inter-city flows (**B**) of each county within a city as a function of its distance to the core county. The percentage of outflows due to intra- and inter-city flows are shown in (**C**) and (**D**), respectively. The positive correlation of the relative growth with distance due to intra-city flows in (**E**), along with the lack of correlation due to inter-city flows in (**F**), indicates that intra-city flows are mainly responsible for increasing the population in the external regions of cities. The sizes of red circles and blue squares are proportional to the city population. The range of distances is split into equally spaced bins. The number of counties n within each bin, from left to right, is 46, 1, 4, 7, 7, 17, 21, 31, 36, 38, 34, 31, 31, 30, 20, 20, 21, 14, 17, 9, 9, 6, 4, 2, 5, 5, 2, 1. The black dots and the error bars indicate the mean and the 90% interval, respectively, of the counties within the corresponding bin. We also show the Pearson correlation coefficient R and the p-value associated with the two-sided test of the null hypothesis of non-correlation.

inflows from other cities is higher in the core county and decays as we move towards the suburbs (Fig. 3B). The moderate negative correlation of this percentage with the distance reveals that inflows from other cities are more likely to concentrate in the core regions of a city.

The percentage of outflows directed from the core county to other counties within the same city has a slightly negative correlation with the distance of the origin county to the city center, so it is more likely to find intra-city flows with outflows from internal regions (Fig. 3C). The core county is an exception again, suggesting that it is less likely that someone leaving the core county will move to another county within the same city. The slightly negative correlation of the percentage of outflows directed to other cities suggests that there is a trend of people leaving the core county and the central regions to move to other cities (Fig. 3D). The high percentage of inflows (Fig. 3B) and outflows (Fig. 3D) in the central region due to inter-city flows implies that the central regions of cities are more dynamic and diverse and that people tend to move to counties with similar levels of urbanization. The same pattern is observed for flows between metro and micro areas, and for metro and non-statistical areas, allowing us to conclude that people moving from rural areas are more likely to move to the external regions of a city (Supplementary Fig. 7).

The positive correlation of the relative growth with the distance due to intra-city flows (Fig. 3E) shows that the resulting intra-city redistribution of people, given by the difference between inflows and outflows, is such that there is a trend from core county to the external counties (viz. suburbs). When compared to the relative growth due to inter-city flows (Fig. 3F), which do not show any trend and that have negative values for the most distant counties, it becomes clear that intra-city flows play a major role in the population increase observed in outer regions of cities. Interestingly, large circle and square dots in Fig. 3E and F suggest that the loss of people due to inter-city netflows is more intense than the gain of people due to intra-city netflows in some external counties of the largest metro areas, thus explaining the population decline in some outer regions of New York and Chicago (as shown in Fig. 2B and C).

The population growth due to intra-city flows is also depicted in Fig. 4. The concentration of flows below the diagonal captures the heterogeneity and the preferential destination of intra-city netflows. We observe that people are more likely to move to lower population density counties when moving from one place to another within the same city, as exemplified by 7 cities in panel A. Panel B summarizes this analysis for the 46 cities with more than 5 counties by showing the fraction $\mathcal{F}$ of intra-city netflows to lower density counties. We note that more than 93% of the cities have $\mathcal{F}$>0.5 and that there is a positive correlation of $\mathcal{F}$ with the city population, and C shows the rank of cities according to the fraction of intra-city netflows to lower density counties.

https://doi.org/10.1038/s41467-022-33527-y

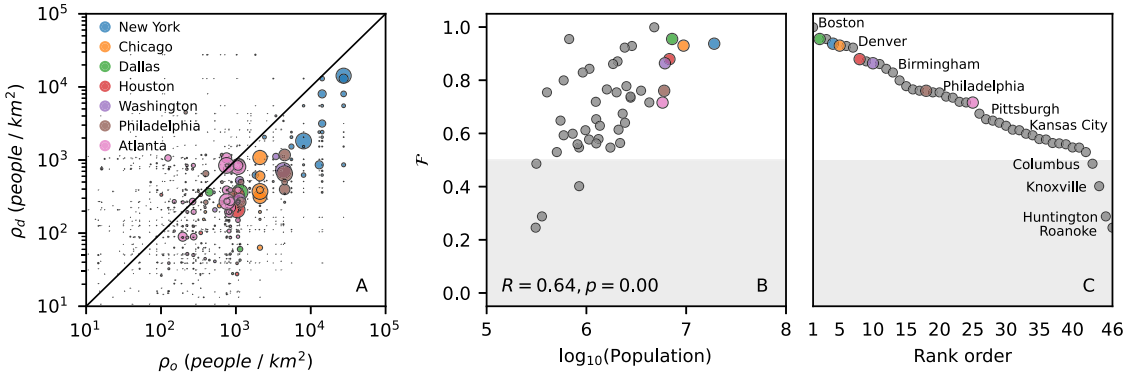

Fig. 4 | People are moving to counties with lower population density. A The population density of the origin (ρ_o) and destination (ρ_d) counties of intra-city netflows for New York, Chicago, Dallas, Houston, Washington D.C., Philadelphia, Atlanta, reveal that the majority of the flows occur from high to low-density counties. The size of the symbols are proportional to the intensity of the netflow, and the black line corresponds to $y = x$. **B** The fraction of netflows to lower density counties $\mathcal{F}$ has a positive correlation with city population when we consider the 46 MSAs with more than 5 counties, suggesting that intra-city netflows to lower density counties are more frequent as the city size increases. We also show the Pearson correlation coefficient R and the p-value associated with the two-sided test of the null hypothesis of non-correlation. **C** The ranking of the cities according to $\mathcal{F}$.

Population density does not seem to play a major role in driving flows between counties of different cities. The fraction of inter-city netflows to lower density counties is about 57% when we consider all the 384 MSAs. The heterogeneity in the inter-city netflow pattern can be assessed by analyzing $\mathcal{F}$ versus the population of the destination city (Fig. 5A, B) and $\mathcal{F}$ versus the population of the origin city (Fig. 5C, D). The negative correlation of $\mathcal{F}$ with the population of the destination city in panel A indicates that inflows are more likely to come from lower density counties as the destination city size increases. The positive correlation of $\mathcal{F}$ with the population of the origin city in panel C reveals that outflows tend to be directed to lower density counties as the origin city size increases. The trends observed in panels A and C reveal that inter-city flows are more likely between counties with different population densities rather than between counties with similar population densities. Panels B and D show the rank order of cities according to a function of the destination city size and the origin city size, respectively.

We would expect that there might be preferential locations within a given city to which people move due to various factors such as lower costs of housing and employment opportunities. However, it seems that house prices have little to no effect on intra-city netflows (Supplementary Fig. 8). While the fraction of intra-city netflows to counties with less expensive houses is about 0.8 for cities like New York, Chicago and Washington, this fraction is about 0.2 for cities like Dallas, Houston and Philadelphia. The lack of a clear national pattern highlights the specificity of each city and the heterogeneity of the regional housing market in the U.S.[36,37]. On the other hand, the fraction of intra-city netflows to counties with lower unemployment rates is higher than 0.5 for about 2/3 of the cities (Supplementary Fig. 9), thus showing that people are more likely to move to counties with lower unemployment rates.

Statistical structure of inter-city flows

Intra-city flows capture the internal redistribution of population, without altering the total city population. In this context, we focus on inter-city flows to investigate whether or not extreme flows play an important role in shaping the growth of counties as observed at the city level[5]. For cities, Verbavatz and Barthelemy[5] introduce a stochastic equation to describe population growth, composed of two terms. The first term accounts for out-of-system growth, which includes natural growth and international migration, and the second term accounts for the growth due to domestic netflows. They find that total netflows adjusted by population size can be well approximated by a Lévy distribution, and this heavy-tailed distribution indicates that rare and extreme inter-city flows (viz. migratory shocks) dominate city population growth.

Here, we find that, for counties, the distribution of total netflows adjusted by population size, which is represented by ζ_i and captures the intensity of inter-city migratory flows (see the section "Methods" for details), can be approximated by a Gaussian distribution (Fig. 6). The lack of a heavy tail in the empirical distribution of ζ_i suggests the absence of extreme flows at the county level, thus indicating that the growth of counties can be described by smoother migratory process than cities. Given that cities do experience migratory shocks[5], our findings indicate that cities redistribute inflows among its different counties, leading to a spill-over effect that dampens flow shocks at the county level.

Heterogeneity of international inflows

The highest share of international inflows is concentrated in large cities. About 40% of the international inflows are destined to the top 10 (~2.6%) largest metro areas of the U.S. New York is the first with 8.5% of international inflows, followed by Los Angeles and Miami with 5.4% and 5.0%, respectively. Indeed, international inflows Y_k scale superlinearly with the population S_k of the metro area k (Fig. 7A), thus larger cities have more immigrants per capita than smaller cities.

Interestingly, this gain with scale is also observed in socioeconomic city metrics as crime, GDP, innovation and wealth creation due to the manifestation of nonlinear agglomeration phenomena[38–40]. Using $Y = Y_0 S^\theta$ as a null model, we can compute deviations from the average behavior by means of residuals given by $\log(Y_k / Y_0 S_k^\theta)$[38]. The rank of the residues (Fig. 7B) shows that college towns are among the top metro areas receiving more international inflows than expected, while large cities as Los Angeles, New York, Atlanta, and Chicago are among the metro areas receiving less international inflows than expected.

The spatial distribution of international inflows within cities is shown in Supplementary Fig. 10. The highest share of inflows is concentrated at core counties, and the percentage of inflows decreases dramatically with the distance from the core county. This result suggests that inflow of international migrants is an important component of population growth, particularly at the core regions of large cities.

Robustness of our findings

Patterns of population redistribution change from time to time in the U.S., and are affected by several factors. For instance, in the 1960s nonmetropolitan counties lost about 3 million people due to outflows to metropolitan counties, while the reverse trend was observed in the

https://doi.org/10.1038/s41467-022-33527-y

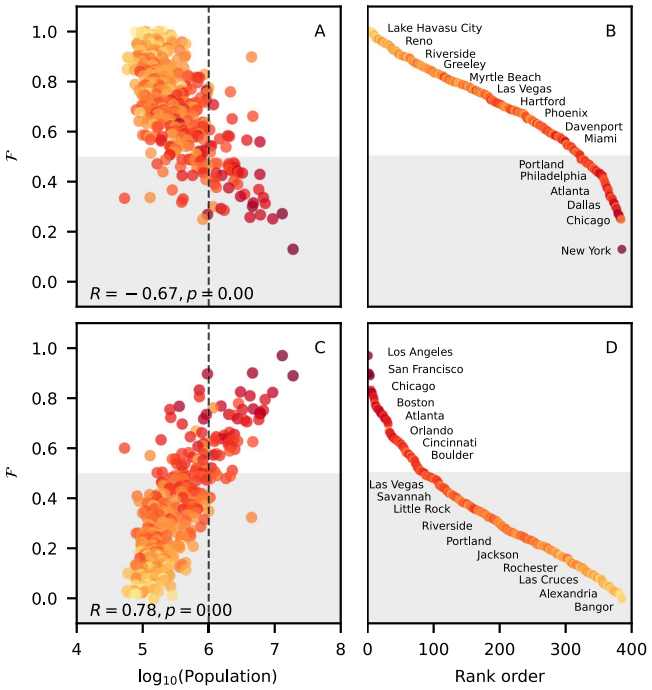

Fig. 5 | Inter-city flow patterns depend on the population size of the origin and destination cities. Each point corresponds to a particular city. **A** Fraction $\mathcal{F}$ of netflows going to lower density counties versus the population of the destination city. Inflows to counties of large cities (with population greater than 10^6, dashed line) usually comes from counties with lower population densities. **B** Rank of cities according to the share of inflows from lower density counties. **C** Fraction $\mathcal{F}$ versus the population of the origin city. Outflows from counties of large cities usually go to cities with lower density counties. **D** The rank of cities according to the share of inter-city netflows to lower density counties is presented. The dots are colored according to the city population density (darker red means higher density). We also show the Pearson correlation coefficient R and the p-value associated with the two-sided test of the null hypothesis of non-correlation.

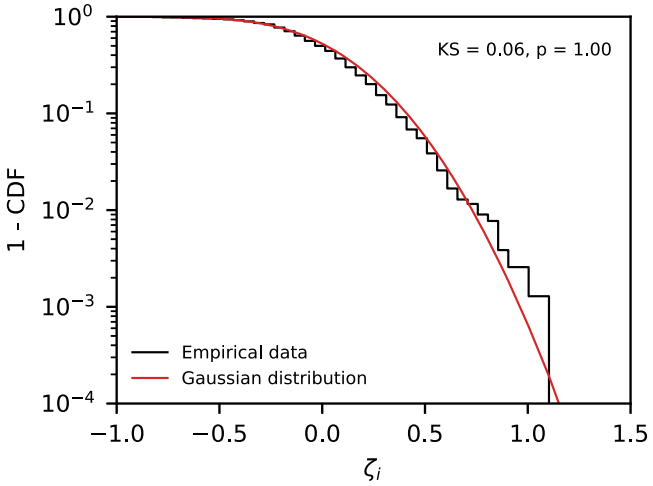

Fig. 6 | Extreme shocks are dissipated at the county level. The distribution of ζ_i, which is the sum of the netflows of a county i adjusted by its population, suggests that migratory events are exponentially bounded at the county level since ζ_i is well described by a Gaussian distribution. The distribution of ζ_i is computed here for all the counties with at least 50.000 inhabitants. We also show the result of the two-sided KS test under the null hypothesis that ζ_i follows a Gaussian distribution.

1970s when non-metropolitan counties experienced net inflows of about 2.6 million people[41]. Wardwell and Brown in[41] indicate that three factors might be among the main reasons of such change, namely economic decentralization, preference for rural living, and modernization of rural life. The temporal influence of factors as socio-economic conditions, transportation infrastructure, natural amenities, and land use and development on population growth in rural and suburban areas is explored in[42]. Changes in rural migration patterns are also studied in[43], where age-specific rural migration patterns from 1950 to 1995 are analyzed. In[44], the authors explore redistribution trends across U.S. counties from 1980 to 1995 split into three five year periods (1980–1985, 1985–1990, 1990–1995), and[45] analyzes changes in age-specific nationwide migration patterns from 1950 to 2010.

The spatial structure of migration patterns may indeed change from time to time; our results correspond to the current intra- and inter-city redistribution trends, based on the most recent ACS migration flow files. We present a thorough empirical and statistical analysis of domestic migration flows among U.S. cities ans counties. Our study also introduces a framework that can be used for analyzing and comparing internal redistribution of people across different time periods. Indeed, we extended our analysis for two other time periods, 2005–2009 and 2010–2014. With respect to the spatial distribution of intra- and inter-city flows, similar trends are observed in both periods (Supplementary Figs. 12, 13), namely inter-city flows are responsible for the highest share of inflows to core counties, and intra-city flows are responsible for the highest share of inflows to external counties. We also explored the role of population density in driving netflows between counties within the same metro area in 2005–2009 and 2010–2014. The results (Supplementary Figs. 14 and 15) indicate that 95.7% of cities were dominated by intra-city moves to lower density counties in 2005–2009, and this percentage dropped to 76.1% in 2010–2014. Our findings indicate that the trends we report here are taking place since 2005 but with different intensities.

The robustness of our findings is checked with additional migration data from the Internal Revenue Service (IRS), which reports the year-to-year address changes on individual tax returns filled with the IRS[46]. The results obtained with the analysis of IRS datasets from periods 2015–2016, 2016–2017, 2017–2018, 2018–2019 (Supplementary Figs. 16, 17, 18, 19), reveal similar trends to those we found using ACS data. Particularly, we observe that, for all periods considered, the correlation between intra-city netflow/S and distance to core county is stronger than we found with ACS data, thus highlighting the role of intra-city flows in driving population to external regions of cities. The main difference between both datasets is in the percentage of intra- and inter-city inflows and outflows: while ACS data indicates that both flows have approximately the same contribution to the total flows, the IRS data indicates that, besides the core county, intra-city flows are responsible for about 80% of inflows and outflows of metro areas.

Discussion

We presented an analysis of the domestic population flows from domestic migration, disaggregated at the county level, that drive the population growth of U.S. counties and cities. We showed that urban population growth is spatially heterogeneous, where intra- and inter-city flows contribute equally to the population dynamics of cities. Intra-county flows could not be examined in this study (absence of sub-county data), but account for ~60% of all domestic flows in the U.S. (Supplementary Fig. 1). Analyzing spatial aspects of these flows is an interesting direction for future studies.

Large polycentric urban agglomerations in the U.S. emerge from the development and merging of several towns and cities with strong socioeconomic ties, even though they retain separate municipal governing structures[47,48]. Similarly, at the regional-scale, even larger urban agglomerations of multiple cities emerge; OMB designated 172 such agglomerations as Combined Statistical Areas. Our analyses of intra-city flows are based on core county, and the spread of data points observed in Fig. 3 reveals the heterogeneity among cities in terms of

https://doi.org/10.1038/s41467-022-33527-y

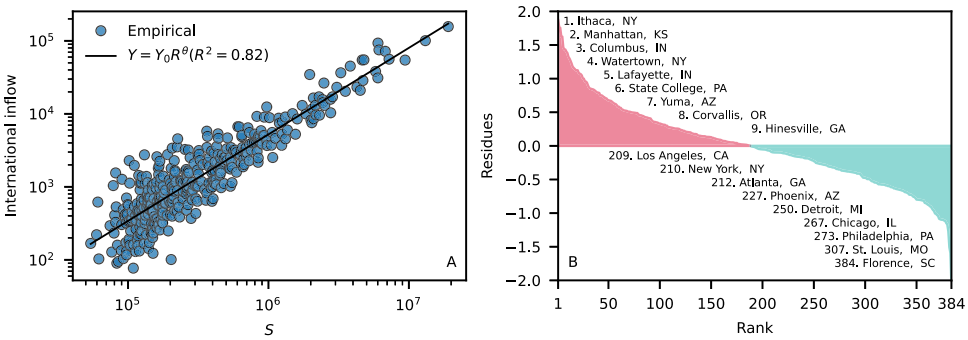

Fig. 7 | International inflow scales superlinearly with city size. Panel (**A**) shows the number of international immigrants as a function of the city size S for the 384 U.S. metro areas. The performance of the model $Y = Y_0 S^\theta$, in which $\theta = 1.19$ (95% CI [1.13, 1.24]) and $Y_0 = 4.10^{-4}$ is a normalization constant, is assessed by the coefficient of determination R^2. Note that the spread of empirical data around the model narrows as the size of the city increases. Panel (**B**) shows the rank of the metro areas and the residues, which captures the deviation from the null model thus highlighting cities receiving more/less than expected international inflows. Names of the cities are followed by two-letter state abbreviations.

polycentric organization, but with a primary, central urban county[47,48]. Recent multi-scale modeling analyses of urban mobility and growth[49,50],[51,52] are noteworthy in combining diverse data sources and theoretical approaches, but there is a need for empirical data analysis at a more disaggregated level[53,54].

Counties with the highest population density in MSAs constitute the core of cities, characterized by intense inter-city inflows and outflows. Although population growth of cities is shaped by inter-city migratory shocks[5], we showed that counties are not subject to the same population dynamics. This result suggests that flow shocks are dispersed among its counties; the spill-over effect thus dampens the shocks at the county level. The population growth of urban counties through densification is driven by creativity, innovation, and technological advances, but also triggers outflows because of increasing cost of living and decreasing quality of life issues[55–58]. Net outflows from the core (most dense) to external counties expand urban sprawl to neighboring counties. Inter-city migrations, and flows from micro and non-statistical areas to metro areas, are more likely between counties with similar levels of urbanization, showing a preferential flow destination.

Not all domestic migration has the same demographic impact. For instance, migration of young people might contribute to a larger natural growth. The migration of elderly people, on the other hand, might have the opposite effect. Particularly, a good discussion about migration up and down urban hierarchy for different age groups is presented in[59], which helps in understanding the trends and the migration patterns we found. Plane et al. show that the main components of positive population growth in higher density urban settings are international immigration and natural increase since domestic netflows are negative. Strong outflows towards lower density counties are composed of people in their late 50s and 60s preferring less congestion, higher natural amenities, and cheaper housing, thus explaining why natural increase is lower at counties far from the core county (Supplementary Fig. 5B). Inflows to higher density counties is mainly composed of young, single, and college-educated adults in the 25-29 year age group. Interestingly, once they reach their mid-career stage and start their family, the migration trends are reversed: we observe trends towards lower density counties for 30-34 and 35-39 year age groups, mainly prompted by housing costs, school quality and suburban road congestion[59].

In this context, we propose a framework for studying the spatial distribution of migratory flows. We have focused here on the U.S. because of the public availability of robust datasets at the county level, but our conclusions could be extended to other developed and highly urbanized countries where the highest share of domestic migration is composed of intra- and inter-city flows. However, the general intra- and inter-city flow patterns we reported here might not hold for low and middle-income countries presenting lower urbanization levels. For instance, city population growth in countries experiencing rapid urbanization in Asia and Sub-Saharan Africa is mainly composed of rural-to-city flows, in which international migration of refugees and the emergence of large informal settlements are also important components of urban growth[60–67].

Metropolitan statistical areas are surrounded by rural counties. As the population of the adjacent regions increase and cities expand, rural counties are reclassified as metro counties. The reclassification of the external counties as urban places in the U.S. fails to recognize that most of the population might in fact be rural[18]. A recent global-scale analysis[68] suggested that about a quarter of the global population lives in periurban areas of intermediate and small cities. Interestingly, they find that the highest share of population is found at large cities and proximate areas for high-income countries, whereas low-income countries have the highest share of population concentrated in small cities and proximate areas. Migration data are needed in order to construct a typology of flows in different parts of the world.

In this paper, we have studied the mechanisms behind the heterogeneous growth of cities. The growth of cities leads to many benefits, which serve to increase the attractiveness of the city. As cities grow, wealth and innovation per capita increases since these quantities scale superlinearly with city size as a result of agglomeration effects[56]. Simultaneously, the volume occupied by infrastructure scales sublinearly with city size, and this economy of scale means that large cities need less infrastructure per capita than small cities[40]. Conversely, land-use changes from the expansion of metro areas also has costs at local, regional and national scales. Such costs include, among others, loss of agricultural areas (food security)[69], fragmentation of natural areas (loss of ecosystem services)[70,71], and growing resource demands extracted from increasingly remote locations (impacts on ecosystem impacts at larger scales)[72]. Heterogeneity of cities both manifests and amplifies socioeconomic inequality, thus contributing diminished urban community resilience.

Understanding the spatial organization of the migratory flows is a step towards understanding the drivers of urban growth and heterogeneities in cities. We found that it is necessary to distinguish these flows into different components according to their destination (central vs. external county) and these flows are probably governed by different underlying reasons and household cost-benefit analyses. The possibility of constructing a typology of flows allows then to test the influence of various parameters and models, to analyze the dynamics of inequalities within cities, and eventually to help urban planners to forecast urban expansion and densification.

Methods

Data collection

Our analysis is focused on a five-year period, from 2015 to 2019. We used two main sources of datasets, both from the U.S. Census. The first main source is the ACS County-to-County Migration Files[26]. Annually, approximately 3.54 million independent housing units addresses were selected among all the U.S. counties. There were four modes of data collection: internet, mail, telephone, and personal visit, in which respondents are asked whether they lived in the same residence one year ago. The results, reported over 5-year periods for robust flow estimates of less populated counties, are made available cleaned and preprocessed in spreadsheet files[26]. From this file, we obtain estimates of inflows and outflows between pairs of counties, in which "flow estimates resemble the annual number of movers between counties for the 5-year period data was collected"[73].

The second main source is the County Population Totals: 2010–2019 dataset[74], which offers "population, population change, and estimated components of population change" from April 1, 2010 to July 1, 2019. Using the resident population from the 2010 Census as a starting point (population base), county population estimates are derived from the following demographic balancing equation: population estimate = population base + births - deaths + migration[74] (see[74] for detailed explanations of how births, deaths and migration are estimated). From this dataset, which is made available cleaned and preprocessed in a spreadsheet table, we obtain the domestic netflows, births and deaths for each county from July 1, 2015 to June 30, 2019 used to compute the quantity x. We focus our study on 3,141 counties and 384 U.S. metro areas.

To analyze the housing prices of origin and destination counties, we used the housing data from Zillow Research[75]. Zillow publishes the Zillow Home Value Index, which reflects the typical values of homes across the U.S. The data are also made available at the county level, cleaned and preprocessed in a spreadsheet table.

The IRS data we used to check the robustness of our findings were collected from the IRS–SOI Tax Stats–Migration Data website[46]. From[46]: "Migration data for the United States are based on year-to-year address changes reported on individual income tax returns filed with the IRS." Data files are made available for download cleaned and preprocessed, in comma-separated values files (.csv file extension).

Analyzing inter-city flows

Here, we show that the distribution of normalized netflows at the county level is exponentially tempered, suggesting that counties do not experience extreme migratory events as do cities. At the county level, we define $J_{i,k}$ as the aggregate flow from county i to all counties of MSA k, and $J_{k,i}$ as the aggregate flow from all the counties of MSA k to county i. Following[5], we assume that $J_{i,k} = I_0 S_i^{\mu} S_k^{\nu} x_{i,k}$, in which I_0 is a constant, S_i is the population of county i, S_k is the total population of MSA k, μ and ν are exponents of S_i and S_k, respectively, and $x_{i,k}$ accounts for random noises and higher order effects. In this notation, the total netflow of county i is given by $\mathcal{J}_i = \sum_{k \in N_i} (J_{i,k} - J_{k,i})$, where N_i is the set of MSAs exchanging people with county i.

In order to reduce the number of free parameters in the expression for $J_{i,k}$, we define the flow per capita $I_{i,k} = J_{i,k}/S_i$. Given that the ratio $I_{k,i}/I_{i,k} = (S_i/S_k)^{\nu-\mu+1}$ can be written as a linear function of S_i/S_k (see Supplementary Fig. 11A), we obtain $\nu = \mu$. Fitting the flow per capita $I_{i,k}$ versus $S_i^{\nu} S_k^{\nu-1}$ gives us $\nu = 0.34$ (95% CI [0.33, 0.35], Supplementary Fig. 11B). As seen in[5], migratory shocks can be captured by the quantity $X_{i,k} = (J_{i,k} - J_{k,i})/I_0 S_i^{\nu}$, which measures the relative magnitude netflows with respect to the county population. Interestingly, the variable $\zeta_i = (1/N_i) \sum_{k \in N_i} X_{i,k} = \mathcal{J}_i / I_0 N_i S_i^{\nu}$, which is the relative impact of the sum of all netflows in county i and captures the intensity of migratory shocks, can be approximated by a Gaussian distribution (Fig. 6).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

We used data from the following sources in our analysis. The county-to-county migration flows dataset is available at[26]. The county population totals dataset is available at[74]. The housing prices dataset is available at[75]. The IRS migraton flow dataset is available at[46].

Code availability

The data were analyzed in Python 3.7.12, using the following libraries: pandas 1.3.4, matplotlib 3.5.1, numpy 1.18.5, statsmodels 0.13.2, sklearn 1.0.2, scipy 1.7.3. Computer codes will be made available upon request.

References

1. Auerbach, F. Das gesetz der bevölkerungskonzentration. *Petermanns Geographische Mitteilungen* **59**, 74–76 (1913).
2. Zipf, G. K. *Human Behavior and the Principle of Least Effort* (Addison-Wesley, 1949).
3. Gabaix, X. Zipf's law for cities: an explanation. *Q. J. Economics* **114**, 739–767 (1999).
4. Gibrat, R. *Les Inégalités Économiques* (Librairie du Recueil Sirey, 1931).
5. Verbavatz, V. & Barthelemy, M. The growth equation of cities. *Nature* **587**, 397–401 (2020).
6. Bettencourt, L. M. & Zünd, D. Demography and the emergence of universal patterns in urban systems. *Nat. Commun.* **11**, 1–9 (2020).
7. Monras, J. Economic shocks and internal migration. CEPR Discussion Paper No. DP12977 (2018).
8. Šveda, M., Madajová, M. & Podolák, P. et al. Behind the differentiation of suburban development in the hinterland of Bratislava, Slovakia. *Sociologicky` časopis/Czech Sociol. Rev.* **52**, 893–925 (2016).
9. Downs, A. *Still Stuck in Traffic: Coping with Peak-hour Traffic Congestion* (Brookings Institution Press, 2005).
10. Hymel, K. Does traffic congestion reduce employment growth? *J. Urban Economics* **65**, 127–135 (2009).
11. Sweet, M. Does traffic congestion slow the economy? *J. Planning Literature* **26**, 391–404 (2011).
12. Darçín, M. Association between air quality and quality of life. *Environ. Sci. Pollution Res.* **21**, 1954–1959 (2014).
13. Portney, K. E. *Taking Sustainable Cities Seriously: Economic Development, the Environment, and Quality of Life in American Cities* (MIT Press, 2013).
14. Musterd, S., Marcińczak, S., Van Ham, M. & Tammaru, T. Socioeconomic segregation in European capital cities. Increasing separation between poor and rich. *Urban Geogr.* **38**, 1062–1083 (2017).
15. Seto, K. C., Fragkias, M., Güneralp, B. & Reilly, M. K. A meta-analysis of global urban land expansion. *PLoS ONE* **6**, 23777 (2011).
16. Wu, K.-y & Zhang, H. Land use dynamics, built-up land expansion patterns, and driving forces analysis of the fast-growing Hangzhou metropolitan area, eastern China (1978–2008). *Appl. Geogr.* **34**, 137–145 (2012).
17. Schneider, A. & Mertes, C. M. Expansion and growth in Chinese cities, 1978–2010. *Environ. Res. Lett.s* **9**, 024008 (2014).
18. Lichter, D. T., Brown, D. L. & Parisi, D. The rural–urban interface: rural and small town growth at the metropolitan fringe. *Population Space Place* **27**, 2415 (2021).
19. Phelps, N. A. Which city? grounding contemporary urban theory. *J. Planning Literature* **36**, 345–357 (2021).
20. Newling, B. E. The spatial variation of urban population densities. *Geogr. Rev.* **59**, 242–252 (1969).

21. Batty, M. *Cities as Fractals: Simulating Growth and Form* (Springer, 1991).
22. Shreevastava, A., Bhalachandran, S., McGrath, G. S., Huber, M. & Rao, P. S. C. Paradoxical impact of sprawling intra-Urban Heat Islets: Reducing mean surface temperatures while enhancing local extremes. *Sci. Rep.* **9**, 1–10 (2019).
23. Hao, T., Sun, R., Tombe, T. & Zhu, X. The effect of migration policy on growth, structural change, and regional inequality in China. *J. Monetary Econom.* **113**, 112–134 (2020).
24. Slater, P. B. *Hubs and Clusters in the Evolving United States Internal Migration Network*. Preprint at https://arxiv.org/abs/0809.2768 (2008).
25. Slater, P. B. A hierarchical regionalization of Japanese prefectures using 1972 interprefectural migration flows. *Regional Studies* **10**, 123–132 (1976).
26. United States Census Bureau: County-to-County Migration Flows: 2015-2019 ACS. https://www.census.gov/data/tables/2019/demo/geographic-mobility/county-to-county-migration-2015-2019.html. Accessed 8 Feb 2022.
27. United States Census Bureau. https://www.census.gov/. Accessed 8 Feb 2022.
28. Koerber, K. Domestic migration flows for states from the 2005 acs. In: *Annual Meeting of the Population Association of America* (2007).
29. Frost, R. *Are Americans Stuck in Place? Declining Residential Mobility in the U. S.* (Harvard University, 2020).
30. U.S. Dept. of Health & Human Services. https://www.hhs.gov/guidance/document/defining-rural-population. Accessed 18 Dec 2010.
31. Stier, A.J. et al. Reply to Huth et al.: Cities are defined by their spatially aggregated socioeconomic networks. *Proc. Natl Acad. Sci.* **119**, e2119313118 (2022).
32. Batty, M. *Cities as Complex Systems: Scaling, Interaction, Networks, Dynamics and Urban Morphologies* (Springer, 2009).
33. Batty, M. & Cheshire, J. *Cities as flows, cities of flows*. (SAGE Publications Sage UK, 2011).
34. Hamilton, B. E., Martin, J. A., Osterman, M. J. Births: provisional data for 2020. https://doi.org/10.15620/cdc:104993 (2021).
35. Johnson, K. M. As births diminish and deaths increase, natural decrease becomes more widespread in rural America. *Rural Sociol.* **85**, 1045–1058 (2020).
36. Moench, E. & Ng, S. *A Hierarchical Factor Analysis Of Us Housing Market Dynamics*. (Oxford University Press Oxford, 2011).
37. Aastveit, K. A. & Anundsen, A. K. Asymmetric effects of monetary policy in regional housing markets. *IEB Working Paper* (2018).
38. Bettencourt, L. M., Lobo, J., Strumsky, D. & West, G. B. Urban scaling and its deviations: revealing the structure of wealth, innovation and crime across cities. *PLoS ONE* **5**, 13541 (2010).
39. Bettencourt, L. & West, G. A unified theory of urban living. *Nature* **467**, 912–913 (2010).
40. Bettencourt, L. M. The origins of scaling in cities. *Science* **340**, 1438–1441 (2013).
41. Brown, D. L. & Wardwell, J. M. *New Directions in Urban–rural Migration: the Population Turnaround in Rural America* (Elsevier, 2013).
42. Chi, G. & Ventura, S. J. Population change and its driving factors in rural, suburban, and urban areas of Wisconsin, USA, 1970–2000. *population* **12**, 13 (2011).
43. Johnson, K. M. & Fuguitt, G. V. Continuity and change in rural migration patterns, 1950–1995. *Rural Sociol.* **65**, 27–49 (2000).
44. Rayer, S. & Brown, D. L. Geographic diversity of inter-county migration in the united states, 1980–1995. *Population Res. Policy Rev.* **20**, 229–252 (2001).
45. Johnson, K. M. & Winkler, R. L. Migration signatures across the decades: net migration by age in US counties, 1950-2010. *Demogr. Res.* **32**, 1065 (2015).
46. SOI Tax Stats—Migration Data. https://www.irs.gov/statistics/soi-tax-stats-migration-data. Accessed 2 August 2022.
47. Louf, R. & Barthelemy, M. Modeling the polycentric transition of cities. *Phys. Rev. Lett.* **111**, 198702 (2013).
48. Sarkar, S., Wu, H. & Levinson, D. M. Measuring polycentricity via network flows, spatial interaction and percolation. *Urban Stud.***57**, 2402–2422 (2020).
49. Li, X. et al. Global urban growth between 1870 and 2100 from integrated high resolution mapped data and urban dynamic modeling. *Commun. Earth Environ.* **2**, 1–10 (2021).
50. Xu, F., Li, Y., Jin, D., Lu, J. & Song, C. Emergence of urban growth patterns from human mobility behavior. *Nat. Comput. Sci.* **1**, 791–800 (2021).
51. Bettencourt, L. M. Urban growth and the emergent statistics of cities. *Sci. Adv.* **6**, 8812 (2020).
52. Barbosa, H. et al. Uncovering the socioeconomic facets of human mobility. *Scientific Reports* **11**, 1–13 (2021).
53. Yabe, T., Rao, P. S. C., Ukkusuri, S. V. & Cutter, S. L. Toward data-driven, dynamical complex systems approaches to disaster resilience. *Proc. Natl Acad. Sci.* **119**, 2111997119 (2022).
54. Yabe, T., Jones, N. K., Rao, P. S. C., Gonzalez, M. C. & Ukkusuri, S. V. Mobile phone location data for disasters: a review from natural hazards and epidemics. *Comput. Environ. Urban Syst.* **94**, 101777 (2022).
55. Fan, C. C. The vertical and horizontal expansions of China's city system. *Urban Geography* **20**, 493–515 (1999).
56. Bettencourt, L. M., Lobo, J., Helbing, D., Kühnert, C. & West, G. B. Growth, innovation, scaling, and the pace of life in cities. *Proc. Natl Acad. Sci.* **104**, 7301–7306 (2007).
57. Graham, S. Luxified skies: How vertical urban housing became an elite preserve. *City* **19**, 618–645 (2015).
58. Wong, K. G. Vertical cities as a solution for land scarcity: the tallest public housing development in Singapore. *Urban Des. Int.* **9**, 17–30 (2004).
59. Plane, D. A., Henrie, C. J. & Perry, M. J. Migration up and down the urban hierarchy and across the life course. *Proc. Natl Acad. Sci.* **102**, 15313–15318 (2005).
60. Potts, D. Debates about A frican urbanisation, migration and economic growth: what can we learn from Zimbabwe and Zambia? *Geographical J.* **182**, 251–264 (2016).
61. Cockx, L., Colen, L. & De Weerdt, J. From corn to popcorn? Urbanization and food consumption in sub-saharan Africa: evidence from rural-urban migrants in Tanzania. LICOS Discussion Paper Series (2017).
62. Saghir, J. & Santoro, J. Urbanization in Sub-Saharan Africa. In: *Meeting Challenges by Bridging Stakeholders* (Center for Strategic & International Studies, 2018).
63. Østby, G. Rural–urban migration, inequality and urban social disorder: evidence from African and Asian cities. *Conflict Manage. Peace Sci.* **33**, 491–515 (2016).
64. Lohnert, B. *Migration and the Rural-urban Transition in Sub-saharan Africa* (Humboldt-Universität zu Berlin, 2017).
65. Zhang, K. H. & Shunfeng, S. Rural–urban migration and urbanization in China: evidence from time-series and cross-section analyses. *China Econ. Rev.* **14**, 386–400 (2003).
66. Malik, N., Asmi, F., Ali, M. & Rahman, M. M. et al. Major factors leading rapid urbanization in China and Pakistan: a comparative study. *J. Social Sci. Stud.* **5**, 148 (2017).
67. Soman, S., Beukes, A., Nederhood, C., Marchio, N. & Bettencourt, L. Worldwide detection of informal settlements via topological analysis of crowdsourced digital maps. *ISPRS Int. J. Geo-Inform.* **9**, 685 (2020).
68. Cattaneo, A., Nelson, A. & McMenomy, T. Global mapping of urban–rural catchment areas reveals unequal access to services. *Proc. Nal Acad. Sci.* **118**, e2011990118 (2021).

69. Bren d'Amour, C. et al. Future urban land expansion and implications for global croplands. *Proc. Natl Acad. Sci.* **114**, 8939–8944 (2017).
70. Li, G. et al. Global impacts of future urban expansion on terrestrial vertebrate diversity. *Nat. Commun.* **13**, 1–12 (2022).
71. Liu, Z., He, C. & Wu, J. The relationship between habitat loss and fragmentation during urbanization: an empirical evaluation from 16 world cities. *PLoS ONE* **11**, 0154613 (2016).
72. Yeh, C.-T. & Huang, S.-L. *Global Urbanization and Demand for Natural Resources* (Springer, 2012).
73. United States Census Bureau: American Community Survey Migration Flows. https://www.census.gov/data/developers/data-sets/acs-migration-flows.html. Accessed 8 Feb 2022.
74. United States Census Bureau: County Population Totals: 2010-2019. https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-total.html. Accessed 8 Feb 2022.
75. Housing Data - Zillow Research. https://www.zillow.com/research/data/. Accessed 31 Jul 2022.

Acknowledgements

P.S.C.R. was supported, in part, by the Lee A. Reith Endowment in the Lyles School Civil Engineering at Purdue University.

Author contributions

S.M.R., P.S.C.R., M.B., S.V.U. designed the study, S.M.R. acquired the data, and S.M.R., P.S.C.R., M.B., S.V.U. analyzed and interpreted the data and wrote the manuscript.

Competing interests

The authors declare no competing interests

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-33527-y.

Correspondence and requests for materials should be addressed to Satish V. Ukkusuri.

Peer review information *Nature Communications* thanks the other anonymous reviewer(s) for their contribution to the peer review of this work.