OPEN

City structure shapes directional resettlement flows in Australia

Bohdan Slavko [✉], Kirill Glavatskiy & Mikhail Prokopenko

Modern urban science views differences in attractiveness of residential suburbs as the main driver of resettlement within a city. In particular, certain suburbs may attract residents due to lower commute costs, and this is believed to lead to compactification of a city, with highly populated central business district and sprawled suburbia. In this paper we assess residential resettlement patterns in Australian capital cities by analyzing the 2011 and 2016 Australian Census data. Rather than explicitly defining a residential attractiveness of each suburb in subjective terms, we introduce and calibrate a model which quantifies the intra-city migration flows in terms of the attractiveness potentials (and their differences), inferring these from the data. We discover that, despite the existence of well-known static agglomeration patterns favouring central districts over the suburbia, the dynamic flows that shape the intra-city migration over the last decade reveal the preference directed away from the central districts with a high density of jobs and population, towards the less populated suburbs on the periphery. Furthermore, we discover that the relocation distance of such resettlement flows plays a vital role, and explains a significant part of the variation in migration flows: the resettlement flow markedly decreases with the relocation distance. Finally, we propose a conjecture that these directional resettlement flows are explained by the cities' structure, with monocentric cities exhibiting outward flows with much higher reluctance to long-distance relocation. This conjecture is verified across the major Australian capitals: both monocentric (Sydney, Melbourne, Brisbane, Adelaide, Perth, Hobart) and polycentric (Darwin and Canberra).

The modern world undergoes rapid urbanisation¹. Growing complexity of cities demands new fundamental understanding and rigorous models of the urban dynamics, including diverse and transient intra-urban migration patterns. At the same time, the increasing availability of large-scale spatiotemporal data (e.g., from censuses or surveys) provides an additional impetus for quantitative modelling of urban development and related dynamics^{2–4}. A consistent modeling framework for settlement formation does not yet exist², despite an abundance of stable empirical patterns^{5–12} and theoretical insights^{13–22} that can potentially improve our understanding of how cities grow and self-organise.

Contemporary cities are shaped by the long process of settlement formation and evolution which involves multiple types of migration and resettlement. In social sciences it is common to investigate population migration in the context of individual motivation and social consequences^{23–25}. However, heterogeneous preferences of individual residents cannot be easily integrated into their aggregate migration flows. The latter is a collective phenomenon which we believe is determined mainly by the spatial city structure². Furthermore, such migration patterns are usually persistent in time. In particular, we observe strong correlation between migration flows in two different periods of 2006–2011 and 2011–2016 (see Fig. 6 in the Supplementary Material). Focusing on a small number of macroscopic parameters (such as migration flow) instead of a large number of microscopic degrees of freedom (individual human choices) is one of the underlying motivations of this study, developed from the perspective of sociophysics^{26–31}.

Although static properties of modern cities have been studied quite extensively, the spatial evolution of cities over time remains a subject of vigorous research. Often the dynamics of evolution is typically assumed to be determined by the current static structure. In particular, the compact structure of modern cities can be explained by the transportation cost minimisation^{18,32}, so that the already dense central business districts attract even more jobs and more residents, while the sparsely populated city's periphery is being diluted. An actual verification of this division is subject to data availability. For example, various statistics of resettlement processes may be traced on the national level using census data, but typically, detailed information is not collected. In particular, the US³³ or Canadian³⁴ census datasets do not contain records of previous places of residence, and instead merely include

Centre for Complex Systems, The University of Sydney, Sydney, NSW, 2006, Australia. ✉e-mail: bohdan.slavko@sydney.edu.au

Figure 1. Australian capital cities: location on the map, population (P), and area (A), according to 2016 Census.

relocation events — as a result, this prevents tracking intra-urban migration flows and studying their influence on the spatial structure of the cities. In contrast, the Australian Census³⁵ does collect such information, which makes it a source of unique data for dynamic analysis. Thus, a detailed analysis of the Australian Census data can provide insights of how the static agglomeration patterns affect the dynamic evolution of a city.

The notion of subjective suburb's attractiveness is often considered in the context of identifying the city structure^{18,20,27,36}. We introduce a model comprising the concept of a *revealed attractiveness* potential that drives intra-urban migration. This potential absorbs all characteristics of a suburb, which indicate to what extent people are statistically attracted to it — this potential hence affects the aggregate relocation flows. These characteristics may include the distance to the centers of business or retail activity^{37,38}, the public transport accessibility³⁹, the quality of local infrastructure (schools and other local services)^{36,40}, the housing affordability^{37,41}, and perhaps, other publicly known but not explicitly defined subjective factors. Our model treats the attractiveness as a revealed quantitative characteristic of a suburb, inferring it from data, rather than being based on a particular subjective characterization.

Using our resettlement model, we analyze intra-urban migration flows in the Australian capital cities, represented by the corresponding Greater Capital Areas (GCA): Sydney, Melbourne, Brisbane, Adelaide, Perth, Hobart, Darwin, and Canberra (Fig. 1). These cities are the major centres of business, economic and cultural activities across Australian states and territories, and are well distanced from each other. Furthermore, these cities vary in trajectories of their historical development, ranging from short periods of centralized planning to long intervals of self-organized evolution. The general diversity observed across major urban centres is expected to be reflected in patterns of modern urban dynamics^{42,43} and, in particular, in patterns of the resettlement dynamics.

We use the latest Census data to estimate the spatial distribution of revealed attractiveness and its contribution to the migration flows. As a result, we discover that the role of the suburbs' attractiveness is opposite to what may be expected by a naive perspective: the migration flows are directed away from the central districts with a high population density and infrastructure accessibility, towards sparsely populated suburbs (typically located at the city's outskirts). In addition, the attractiveness of suburbs alone does not allow one to predict the resettlement flows with an adequate precision. In particular, we observe that, to a large extent, the migration flows are determined by the relocation distance: the short-distance relocations occur much more often than the long-distance ones.

Finally, we observe and quantify a specific relationship between the dynamic migration patterns and the static city structure, defined in terms of its polycentricity (as illustrated by e.g. Fig. 2). Namely, six out of eight capital cities (Sydney, Melbourne, Brisbane, Adelaide, Perth, Hobart) are monocentric, and at the same time exhibit a high *reluctance to long-distance relocations*, with a clear center-to-periphery migration flow. In contrast, two out of eight capital cities (Darwin and Canberra) are non-monocentric, while exhibiting a low reluctance to long-distance relocations without a well-defined flow direction. While it is hard to make a definite conclusion based on small number of cities, we make a conjecture: not only does the human resettlement affect the spatial structure of cities, but the city structure itself may also affect the intra-urban migration. A proper verification of this conjecture is subject to future research and data availability.

Results

We analyze residential migration using three related hypotheses, which are referred here as the baseline hypothesis, the extended hypothesis, and the spatial hypothesis. The baseline hypothesis investigates the sole effect of the attractiveness potentials *per se*. The extended hypothesis augments the baseline with the relocation distance dependence. Finally, the spatial hypothesis focuses on the relocation distance alone.

According to all hypotheses the migration flow T_{ik} (amount of relocated people per unit of time) from suburb i to suburb k is proportional to the populations P_i and P_k of either suburb, as well as to the migration driving force $\exp(f_{ik})$:

Figure 2. Urban sprawl of GCA, according to 2016 Census. Yellow color corresponds to the lowest population, while blue color corresponds to the highest population. The bar on the right encodes the value of the density measured in the amount of people per km². Note that the density range and therefore the color correspondence are different for each GCA. The suburbs with the population density less than 10 people/km² are not shown.

$$T_{ik} = \mu P_i P_k e^{f_{ik}}, \quad (1)$$

where μ is a positive normalizing constant. In particular, we fit Census data on migration flows T_{ik} to the model equations and judge on the validity of each hypothesis based on the goodness of fit and on the number of fitting parameters, which are summarized in Table 1.

GCA	# of parameters			R^2		
	baseline	extended	spatial	baseline	extended	spatial
Sydney	313	314	2	0.04	0.68	0.59
Melbourne	310	311	2	0.04	0.70	0.58
Brisbane	237	238	2	0.06	0.68	0.64
Adelaide	111	112	2	0.13	0.76	0.72
Perth	174	175	2	0.13	0.72	0.58
Hobart	36	37	2	0.16	0.65	0.64
Darwin	45	46	2	0.36	0.62	0.45
Canberra	132	133	2	0.36	0.78	0.62

Table 1. Calibration details of migration models (2)–(4) fitted to the 2011–2016 data.

We observe, that the baseline hypothesis (2) provides the worst fit, both in terms of the coefficient of determination and the number of parameters. In contrast, the extended hypothesis (3) uses only one parameter more, but provides the best fit in terms of the coefficient of determinations. Finally, the spatial hypothesis (4) isolates the differences between the extended and the baseline hypothesis and uses 2 parameters only, providing the goodness of fit comparable to the extended hypothesis. This indicates that the spatial hypothesis captures the main pattern in relocations, while the extended hypothesis provides additional insights. Below we describe each of these hypotheses in detail.

Baseline migration hypothesis: attractiveness potentials. To assess the impact of suburb attractiveness on residential migration flows, we adopt the migration model of Weidlich and Haag^{15,40} which is based on the analogy between the processes of human resettlement and chemical reactions. For each suburb i we calculate a value of the potential u_i , which measures the suburb's attractiveness. The higher the (positive) value of the potential is, the more attractive the suburb is. It can be interpreted as an aggregated indicator which combines all possible characteristics attributed to an individual suburb i . These characteristics may include population density, distance to centers of business and retail activity, quality of public services (schools, parks, etc.), transportation connectivity, and other hedonic indicators of the suburb. The notion of the suburb's potential was introduced in the analysis of inter-regional migration⁴⁴, and its role in the dynamics of residential resettlement has been investigated in our earlier study⁴⁰. In particular, we demonstrated, that the model incorporating the suburb's potentials contains critical regimes where small fluctuations in the model parameters may result in sharp changes in the spatial population structure. This aspect of the approach is comparable to the Boltzmann-Lotka-Volterra model (also known as Harris and Wilson model) which describes an evolution of the urban structures^{13,27}, predicting a qualitatively similar long-term outcome⁴⁰.

In the baseline hypothesis, the migration force f_{ik} from suburb i to suburb k is modelled as

$$f_{ik} = u_k - u_i. \quad (2)$$

where the sign of the migration force is chosen in such a way that the migration flows are directed from a suburb with the lower potential to a suburb with the higher potential. Instead of assuming how the suburb's potential depends on its characteristics, we *infer their values directly from the data*. In particular, the suburb's potentials u_i are treated as independent parameters. For a city with N suburbs there are $N+1$ parameters which are fitted to $N(N-1)$ migration flows between each pair of suburbs. This model allows us to test the hypothesis that the migration flow from suburb i to suburb k depends only on the difference in individual characteristics of these suburbs (which we call “attractiveness”), and does not depend on other pairwise characteristics (such as distance or other types of proximity). In other words, we test whether the attractiveness u_k of suburb k equally enhances the migration flows T_{ik} from all suburbs i , regardless of their proximity to k .

The results of calibration for the period of 2011–2016, and the corresponding number of model parameters are presented in Table [1] (and in Table [3] in the supplementary material for the 2006–2011; the revealed values of attractiveness are displayed in Fig. 7 in the Supplementary Material). The determination coefficient R^2 is very low, indicating that the attractiveness potentials u_i fail to explain systematic migration patterns. This in turn means that the impact of any suburb's characteristic on the residential migration is low and the attractiveness alone does not provide significant information on the resettlement flows.

Extended migration hypothesis: attractiveness potentials together with relocation distance.

We next extend the baseline hypothesis to include additional characteristics which may affect the relocation process. In particular, we assume that the force driving a migration flow from suburb i to suburb k depends also on the distance d_{ik} between these suburbs, so that the migration force f_{ik} is set as follows:

$$f_{ik} = u_k - u_i - \gamma d_{ik}, \quad (3)$$

where γ is the same for all suburbs. This model extends the baseline model (2) by adding a pairwise characteristic d_{ik} , i.e., the distance between the destination k and the source i of the corresponding migration flow T_{ik} . In particular, the migration flow to suburb k is determined not only by its intrinsic attractiveness, but also by the proximity of a source i to the destination k .

GCA	baseline	extended	spatial
Sydney	0	1.01	0.91
Melbourne	0	0.94	0.88
Brisbane	0	0.83	0.79
Adelaide	0	0.82	0.78
Perth	0	0.85	0.76
Hobart	0	0.74	0.73
Darwin	0	0.25	0.24
Canberra	0	0.36	0.36

Table 2. Relocation impedance calibrated from the 2011–2016 data using models (3) and (4). Model (2) essentially implies $\gamma = 0$.

The factor γ can be interpreted as the dissatisfaction associated with the relocation distance, and will be referred to as the relocation impedance. A high value of γ means that people relocate mostly to neighbouring suburbs, while a lower value of γ means that people relocate both to neighbouring and more distant suburbs with equal preference.

Distance-dependent migration is investigated in various areas of social and biological sciences, such as population biology^{45–47}, daily commutes⁶, shopping activity^{5,48}, trafficking of goods^{6,49,50}, inter-city phone calls⁵¹. There are various functional forms of the distance dependence. The most common ones are the inverse quadratic^{48,50,51}, exponential^{13,36,52,53}, their combination⁵⁰ and Gaussian-like^{46,54} functions. We adopt this approach to describe the internal city migration, and use the exponential form of the distance dependence.

The results of the calibration include now the value of the relocation impedance, which is listed in Table 2, in addition to the values of the attractiveness of each suburb (Fig. 3). The quality of fit (Table 1) of the extended hypothesis increases substantially in comparison to the baseline hypothesis (2). The values of the coefficient of determination R^2 indicate that contribution of the individual potentials to the quality of the overall fit is much lower than the contribution of the distance between the suburbs. This means, among other things, that the share of relocations which are driven by the attractiveness is much smaller in comparison to the relocations driven by geographic proximity.

In addition, we observe that the fitted values of the relocation impedance γ reveal a certain pattern. For the majority of the cities the values of γ are within the interval 0.75–0.90. In contrast, the relocation impedance values for Canberra and Darwin are within 0.25–0.35, i.e., the range which significantly differs from other capital cities.

Finally, comparing the values of the attractiveness potentials in Fig. 3 with the population density in Fig. 2, we observe a notable correlation between them (as shown in Fig. 8 in the Supplementary Material). In the majority of the cities the densely populated central suburbs appear to be less attractive for residential migration, while the remote suburbs seem to attract a higher number of new residents. In contrast, in Darwin and Canberra, the migration flows are directed to the suburbs around the highly populated areas. These observations are consistent with theoretical studies^{15,40} where it is assumed that the attractiveness potential increases with the population density only until a certain threshold level, beyond which the attractiveness goes down. This means, in particular, that for the majority of Australian capitals this threshold value is quite low. This is fairly surprising, since even the largest Australian capitals have very low population density compared with European and Asian cities which have not reached such threshold yet (cf. 20,781 people/km² in Paris⁵⁵ and 6,158 people/km² in Tokyo⁵⁶ vs the values in Table 4 in the supplementary material in Australian capitals).

It is worth noting that this pattern becomes visible only in the extended model (3), whereas the values of potentials u_i , estimated in accordance with the baseline model (2) are not well defined (Fig. 7 in the Supplementary Material). This demonstrates the importance of including the relocation distance into the migration model for a sensible estimation of u_i .

Spatial migration hypothesis: relocation distance without attractiveness potentials. In order to contrast the roles of the relocation distance and the attractiveness potentials, we may consider a simplification of the extended hypothesis, by assuming that the migration force f_{ik} is determined by only the relocation distance:

$$f_{ik} = -\gamma d_{ik}. \quad (4)$$

In this way we abstract away a large number of parameters (the attractiveness of each individual suburb), and keep only two parameters: the relocation impedance and the normalization factor.

The calibrated values of the relocation distance are shown in Table 2, while the quality of fit is shown in Table 1 (and Table 3 in the supplementary material). The calibrated values of γ for the spatial model are similar to the ones for the extended model (3). Canberra and Darwin substantially differ from the other capitals by the values of relocation impedance having magnitude in the range of 0.22–0.24 (Darwin) and 0.34–0.36 (Canberra), compared to 0.73–0.94 in the other cities. Despite a significant reduction in the number of model parameters (from “number of suburbs” plus two in (3) to just two in (4)), the goodness of fit has not changed substantially (R^2 decreased by 10–15 percentage points).

Figure 3. Values of the revealed attractiveness u_i fitted to 2011–2016 migration data. Potential of each suburb varies from the highest value (blue) to the lowest value (yellow) for each particular GCA. The bar on the right encodes the value of the potentials u_i , whose values are normalized so that minimum value is zero. Note that the density range and therefore the color correspondence are different for each GCA. The suburbs with the population density less than 10 people/km² are not shown.

Relating migration patterns to spatial structure of cities. The presented results lead to two important conclusions. Firstly, we observe that temporal stability in intra-urban migration flows (Fig. 6 in the Supplementary Material) can be explained by the fact that the most intensive migration occur between suburbs that are geographically close, while the individual characteristics of suburbs have a significantly smaller influence

Figure 4. Resettlement diagram for Australian capital cities. The vertical axis measures the spreading index (5) at a threshold population density (Table 4 in the supplementary material), while the horizontal axis measures the city's relocation impedance of the spatial model ((4), Table 2). Each GCA corresponds to a point on the diagram.

on the residential migration in the majority of GCA. Secondly, there are two capital cities whose migration patterns substantially differ from those shown by the other capitals. In particular, in Darwin and Canberra, the suburbs' characteristics (represented by the attractiveness potentials) play more important role in shaping the residential migration, and the long distance relocations occur more often than in other capital cities. In this section we investigate how the spatial city structure affects these relocation patterns.

First, we note that canonical properties used to characterize cities, such as the population or the area (Fig. 1), do not explain the observed difference in the migration patterns, as evidenced by Figs. 10 and 11 in the supplementary materials. In particular, Hobart has low population and area but displays a high relocation impedance. This contrast is also observed in small non-capital cities with a sufficiently representative number of suburbs. Furthermore, the overall variation in the population and area of the capital cities does not concur with the variation of the relocation impedance. This indicates that the relocation impedance may depend on other, more refined, characteristics of a city which better reflect its internal structure.

There exists a number of approaches which identify the most important characteristics of urban spatial structures, including compactness and spreading^{10,57}, spatial clustering of hotspots (areas with high population density or intensive economic activity)⁵⁸, monocentricity versus polycentricity⁵⁹, population density profile⁶⁰, etc. In a recent study¹⁰, two main indicators capturing the whole diversity of the possible urban structures have been introduced: the degree of heterogeneity and the degree of spreading. Here we adopt the latter, and investigate to what extent the suburbs with a high population density are localized together, and how far they are spread across a city.

The spreading index¹⁰ is defined as the ratio of the average distance between the highly populated areas to the average distance between all areas in a city:

$$\eta(\rho) = \frac{\sum_{i,j} d_{ij} \xi_i(\rho) \xi_j(\rho)}{\sum_{i,j} d_{ij} S_i S_j} \quad (5)$$

where $\xi_i(\rho) \equiv S_i \Theta(\rho_i - \rho) / S(\rho)$, with $\Theta(x)$ being the Heaviside step function, S_i is the area of suburb i , S is the total area of suburbs, and $S(\rho)$ is the area of suburbs with the population density higher than ρ . The spreading index is a function of the density, declining rapidly from 1 to 0 if the city is monocentric, in which case the most populated suburbs are heavily clustered within a single area. In contrast, if the suburbs with high population density are spread all around the city, $\eta(\rho)$ remains high within a large range of ρ values. It is possible to evaluate a single value of the spreading index η^* (instead of the entire function $\eta(\rho)$), for a particular value of the threshold density ρ^* , reflecting the intuitive concepts of city "spreadness" and "monocentricity"^{10,61}. To evaluate particular threshold values of the density ρ^* for each GCA, as well as the spreading index $\eta^* = \eta(\rho^*)$, we follow methods quantifying spatial urban organization with respect to the population density^{10,61}. The threshold densities for each GCA are listed in Table 4 of the supplementary material.

The spreading index η^* is plotted against the corresponding relocation impedance γ in Fig. 4. In cities with a low spreading index we observe a higher relocation impedance (which is typical for compact monocentric cities¹⁰). Conversely, a low relocation impedance is observed in cities with high η^* (which is typical for spread-out polycentric cities¹⁰). We can identify the cluster of monocentric cities with a high relocation impedance, for which γ is in the interval 0.8 ± 0.12 , while η^* is in the interval 0.4 ± 0.14 . This cluster is robust with respect to an extension to non-capital cities with comparable population and the number of suburbs. In particular, the smallest number of suburbs among all GCAs is 35 (Hobart).

Most of non-capital Australian cities have less than 30 suburbs, with only two exceptions (Gold Coast and Sunshine Coast). The relocation impedance values for Gold Coast and Sunshine Coast are 0.68 and 0.81 respectively (not shown in the figure), and the corresponding values of spreading index are 0.51 and 0.41. This means that these two cities belong to the monocentric cluster with the high relocation impedance as well.

Figure 5. Binary density map of GCA for 2016 Census. The population density of each suburb is compared to the threshold density calculated by the LouBar methodology^{10,61} (see Table 4 in the Supplementary Material) of the corresponding city, and is identified as being high (blue) or low (yellow). The suburbs with the population density less than 10 people/km² are not shown.

The spreading index analysis is illustrated in Fig. 5: the population density of each suburb is compared to the threshold density calculated by the LouBar methodology^{10,61} (see Table 4 in the supplementary material) of the corresponding city, and is identified as being high (blue) or low (yellow). We observe that a clear majority of the cities (Sydney, Melbourne, Brisbane, Adelaide, Perth, Hobart) include high-density suburbs, predominantly clustered around a single center. In contrast, the two remaining cities (Darwin and Canberra) include high-density suburbs clustered around multiple centers. Comparing Fig. 5 with the actual density distribution within each

GCA in Fig. 2, we conclude that the adopted method (comprising the binary threshold ρ^* and spreading index η^*) indeed reflects the intuitive perception of a city “spreadness”.

The spreading index reflects structural properties of a city, while the relocation impedance reflects its dynamic properties. The observed correlation between these two quantities suggests that the population dynamics is likely to be influenced by urban spatial structure.

Conclusions

Our analysis of intra-urban resettlement in the Australian Capital Cities leads to several important conclusions. In particular, we observe that the revealed attractiveness of individual suburbs (which may include, among other characteristics, the distance to the city center, population density, quality of infrastructure, etc.) alone cannot explain the migration flows with a satisfactory precision. In contrast, we discover that the relocation distance alone provides much higher determination of migration flows than the suburbs’ attractiveness.

It is interesting to note that while the importance of the geographical distance has been shown for various types of migration (in particular, inter-city migration^{6,62}, intra-urban commutes^{13,52}, or even migration of animals^{63–65}, which obey the so-called “gravity law”), it is generally believed that the relocation distance is not important for intra-urban resettlement. In particular, the traditional understanding of how the distance affects migration relies on the associated transportation costs⁶², which are supposed to be negligible for relocations within the same city. Here we show that this is not the case for resettlement in Australian capital cities, and that the relocation distance alone can explain circa 50% of the variation in migration flows. Moreover, we observe that all capital cities except for Darwin and Canberra form a cluster of monocentric cities with high relocation impedance. In particular, we see that with respect to intra-urban resettlement, the cities which differ in size, population, transportation network and other dynamical details, reveal similar migration patterns which are independent of these features.

In addition, we observe that the spatial structure of the attractiveness revealed from the migration data is opposite to the static settlement patterns with central districts being more popular than the city outskirts. Instead we show that the most attractive suburbs are located not in the central areas with high population density and intensive business and retail activity, but in sparsely populated outskirts. These results are also consistent with observations typically reported by various real estate analysts⁶⁶.

We report two exceptions to these more general observations: Darwin and Canberra. In particular, in these GCAs i) the migration flows are directed towards more densely populated areas; ii) the relocation impedance is low; iii) the spreading index is high. This indicates that the migration dynamics in these cities is significantly different from the one across all other GCAs. While this can be simply a coincidental feature, we speculate that this is a consequence of historical development of these cities, in which centralized governmental planning played a much larger role than in the other GCAs which followed more than 100 years of essentially self-organized evolution. Indeed, a large part of Canberra was built in the period during 1961–1981⁶⁷, in accordance with government housing and developing programs⁶⁸, while Darwin was completely rebuilt twice: after WWII, and in late 1970s following Cyclon Tracy which destroyed 80% of houses^{69,70}. We believe that this can be an indication of the differences between self-organized and planned development trajectories of cities. It would be interesting to verify such a conjecture for other cities which were highly influenced by centralized planning (e.g., Brasilia, Cancun, Singapore, Washington D.C., etc.).

Materials and Methods

To perform our analysis, we use the data about internal migration in eight Australian capital cities (Sydney, Melbourne, Brisbane, Adelaide, Perth, Hobart, Darwin and Canberra) from the Australian Census datasets (2011 and 2016), provided by the Australian Bureau of Statistics³⁵. We analyze the Greater Capital Areas (GCAs), which represent urbanized contiguous areas around the corresponding cities. The Census data are organized by the so-called “statistical areas” (SAs), where people reside. The data also contain information on where people resided at the Census date and 5 years before the Census date. The information is provided on the aggregate level in terms of the number of people, and not on the individual level. The data are available in various resolutions, and we use the “Statistical Areas Level 2” (SA2s) resolution. There are 2,310 SA2 regions in entire Australia and between 30–300 SA2s in each GCA. Each SA2 has a population generally within a range of 3,000–25,000 persons³⁵, while some of SA2s within each GCA, such as national parks or industrial objects, have negligible population. In our analysis, we discard suburbs with the population density less than 10 persons/km², setting the difference between the urban and the rural suburbs. This particular value is chosen to indicate a threshold beyond which the spreading index η^* becomes insensitive to the threshold variation, as shown in Fig. 9 of Supplementary Material.

We built the dataset of migration flows, which makes a correspondence between each pair of statistical areas (i, k) and the amount of people relocated from i to k . The values of d_{ik} are calculated as distances between centroids of suburbs i and k . To take into account the fact that the sizes of suburbs are not uniform (and hence, distances between big suburbs are larger), we normalize d_{ik} by factor $(S_i S_k)^{-1/4}$ where S_i and S_k are the areas of suburb i and k respectively. This means that d_{ik} is a dimensionless quantity, measured in the number of suburbs (as linear measurements of a suburb are proportional to the square root of its size).

In our analysis we do not consider people who do not change their place of residence within the observation period (the case when $i = k$) because their number is significantly larger than the number of those who do, and it might distort the model parameters’ estimates. Moreover, it is reasonable to expect that quantities T_{ii} cannot be described by the same model because the magnitude of migration flow T_{ik} ($i \neq k$) goes up as the time of observation increases, while the number of people T_{ii} staying in the same area within this period goes down. Also, we do not analyze external migration flows (in and out of GCA).

The fitting procedure is performed numerically using the ordinary least square technique and the trust region reflective algorithm implemented in SciPy 1.3.0 library in Python⁷¹ (method `scipy.optimize.leastsq`): $\theta = \operatorname{argmin}_{\theta} \sum_{i,j} (T_{ij} - \hat{T}_{ij}(\theta))^2$, where θ is a vector of parameters that include μ and potentials u_i in model (2), μ , γ and potentials u_i in model (3), μ and γ in model (4), where T_{ij} is actual migration flow from suburb i to suburb j , and $\hat{T}_{ij}(\theta)$ is the migration flow calculated according to the corresponding model for the parameters given by vector θ . Note that despite the form suggested by Eq. (1), minimizing the logarithmic residual is not feasible, as there exist many suburb pairs with zero migration between them.

Data availability

All data needed to evaluate the conclusions in the paper are available from Australian Bureau of Statistics³⁵.

Received: 14 February 2020; Accepted: 27 April 2020;

Published online: 19 May 2020

References

1. United Nations. World urbanization prospects: The 2014 revision, highlights. department of economic and social affairs. *Popul. Div. United Nations* **32** (2014).
2. Barthelemy, M. *The structure and dynamics of cities* (Cambridge University Press, 2016).
3. Zachreson, C. *et al.* Urbanization affects peak timing, prevalence, and bimodality of influenza pandemics in Australia: Results of a census-calibrated model. *Sci. Adv.* **4**, eaau5294 (2018).
4. Glavatskiy, K. S., Prokopenko, M., Carro, A., Ormerod, P. & Harre, M. Explaining herding and volatility in the cyclical price dynamics of urban housing markets using a large scale agent-based model. *arXiv preprint arXiv* **2004**, 07571 (2020).
5. Reilly, W. J. *et al. Methods for the study of retail relationships* (University of Texas at Austin, 1929).
6. Zipf, G. K. The p 1 p 2/d hypothesis: on the intercity movement of persons. *Am. sociological review* **11**, 677–686 (1946).
7. Batty, M. Rank clocks. *Nature* **444**, 592 (2006).
8. Brockmann, D., Hufnagel, L. & Geisel, T. The scaling laws of human travel. *Nature* **439**, 462–465 (2006).
9. Arcaute, E. *et al.* Cities and regions in britain through hierarchical percolation. *Royal Soc. open science* **3**, 150691 (2016).
10. Volpati, V. & Barthelemy, M. The spatial organization of the population density in cities. *arXiv preprint arXiv* **1804**, 00855 (2018).
11. Barbosa, H. *et al.* Human mobility: Models and applications. *Phys. Reports* **734**, 1–74 (2018).
12. Fletcher, R. J. Trajectories of low-density settlements past and present. *Front. Digit. Humanit.* **6**, 14 (2019).
13. Harris, B. & Wilson, A. G. Equilibrium values and dynamics of attractiveness terms in production-constrained spatial interaction models. *Environ. planning A* **10**, 371–388 (1978).
14. Fujita, M. & Ogawa, H. Multiple equilibria and structural transition of non-monocentric urban configurations. *Reg. science urban economics* **12**, 161–196 (1982).
15. Haag, G. & Weidlich, W. A stochastic theory of interregional migration. *Geogr. Analysis* **16**, 331–357 (1984).
16. Weidlich, W. & Munz, M. Settlement formation. *The Annals Reg. Sci.* **24**, 83–106 (1990).
17. Simini, F., González, M. C., Maritan, A. & Barabási, A.-L. A universal model for mobility and migration patterns. *Nature* **484**, 96 (2012).
18. Louf, R. & Barthelemy, M. Modeling the polycentric transition of cities. *Phys. review letters* **111**, 198702 (2013).
19. Osawa, M., Akamatsu, T. & Takayama, Y. Harris and Wilson (1978) model revisited: The spatial period-doubling cascade in an urban retail model. *J. Reg. Sci.* **57**, 442–466 (2017).
20. Ellam, L., Girolami, M., Pavliotis, G. A. & Wilson, A. Stochastic modelling of urban structure. *Proc. Royal Soc. A: Math. Phys. Eng. Sci.* **474**, 20170700 (2018).
21. Carra, G. & Barthelemy, M. A fundamental diagram of urbanization. *Environ. Plan. B: Urban Anal. City Sci.* **46**, 690–706 (2019).
22. Sahasranaman, A. & Bettencourt, L. M. Urban geography and scaling of contemporary indian cities. *J. Royal Soc. Interface* **16**, 20180758 (2019).
23. Rossi, P. H. *Why families move: A study in the social psychology of urban residential mobility* (Free Press, 1955).
24. Hanson, S. & Pratt, G. *Gender, work and space* (Routledge, 2003).
25. Hanson, S. Perspectives on the geographic stability and mobility of people in cities. *Proc. Natl. Acad. Sci.* **102**, 15301–15306 (2005).
26. Helbing, D. & Molnar, P. Social force model for pedestrian dynamics. *Phys. review E* **51**, 4282 (1995).
27. Wilson, A., Boltzmann, Lotka and Volterra and spatial structural evolution: an integrated methodology for some dynamical systems. *J. The Royal Soc. Interface* **5**, 865–871 (2008).
28. Helbing, D. & Yu, W. The outbreak of cooperation among success-driven individuals under noisy conditions. *Proc. Natl. Acad. Sci.* **106**, 3680–3685 (2009).
29. Bouchaud, J.-P. Crises and collective socio-economic phenomena: simple models and challenges. *J. Stat. Phys.* **151**, 567–606 (2013).
30. Hernando, A., Hernando, R., Plastino, A. & Plastino, A. The workings of the maximum entropy principle in collective human behaviour. *J. The Royal Soc. Interface* **10**, 20120758 (2013).
31. Harding, N., Nigmatullin, R. & Prokopenko, M. Thermodynamic efficiency of contagions: a statistical mechanical analysis of the sis epidemic model. *Interface focus* **8**, 20180036 (2018).
32. Fujita, M. The evolution of spatial economics: from thünen to the new economic geography. *The Jpn. Econ. Rev.* **61**, 1–32 (2010).
33. United States Census Bureau. American fact finder, <https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml>.
34. Statistics Canada. 2016 census profile web data service, <https://www12.statcan.gc.ca/wds-sdw/cpr2016-eng.cfm>.
35. Australian Bureau of Statistics. TableBuilder. <http://www.abs.gov.au/websitedbs/D3310114.nsf/Home/2016%20TableBuilder/>.
36. Crosato, E., Nigmatullin, R. & Prokopenko, M. On critical dynamics and thermodynamic efficiency of urban transformations. *Royal Soc. open science* **5**, 180863 (2018).
37. Kim, J. H., Pagliara, F. & Preston, J. The intention to move and residential location choice behaviour. *Urban studies* **42**, 1621–1636 (2005).
38. Pérez, P. E. & Martínez, F. J. & Ortúzar, J. d. D. Microeconomic formulation and estimation of a residential location choice model: implications for the value of time. *J. regional science* **43**, 771–789 (2003).
39. Wu, H., Levinson, D. & Sarkar, S. How transit scaling shapes cities. *Nat. Sustain.* **2**, 1142–1148 (2019).
40. Slavko, B., Glavatskiy, K. & Prokopenko, M. Dynamic resettlement as a mechanism of phase transitions in urban configurations. *Phys. Rev. E* **99**, 042143 (2019).
41. McFadden, D. Modeling the choice of residential location. *Transp. Res. Rec.* **673** (1978).
42. Barthelemy, M., Bordin, P., Berestycki, H. & Gribaudi, M. Self-organization versus top-down planning in the evolution of a city. *Sci. reports* **3**, 2153 (2013).
43. Bertaud, A. & Malpezzi, S. The spatial distribution of population in 35 world cities: the role of markets, planning and topography. *The Cent. for urban land economic research, The Univ. Wis.* (1999).
44. Weidlich, W. & Haag, G. *Interregional migration: dynamic theory and comparative analysis*, vol. 4 (Springer, 1988).

45. Kareiva, P. Local movement in herbivorous insects: applying a passive diffusion model to mark-recapture field experiments. *Oecologia* **57**, 322–327 (1983).
46. Okubo, A. Dynamical aspects of animal grouping: swarms, schools, flocks, and herds. *Adv. biophysics* **22**, 1–94 (1986).
47. Jopp, F. Comparative studies on the dispersal of the great ramshorn (planorbis corneus l.): A modelling approach. *Limnologia* **36**, 17–25 (2006).
48. Huff, D. L. Defining and estimating a trading area. *The J. Mark.* 34–38 (1964).
49. Jung, W.-S., Wang, F. & Stanley, H. E. Gravity model in the korean highway. *EPL (Europhysics Lett.)* **81**, 48005 (2008).
50. Kaluza, P., Kölsch, A., Gastner, M. T. & Blasius, B. The complex network of global cargo ship movements. *J. Royal Soc. Interface* **7**, 1093–1103 (2010).
51. Krings, G., Calabrese, F., Ratti, C. & Blondel, V. D. Urban gravity: a model for inter-city telecommunication flows. *J. Stat. Mech. Theory Exp.* **2009**, L07003 (2009).
52. Wilson, A. G. The use of entropy maximising models in the theory of trip distribution, mode split and route split. *J. Transp. Econ. Policy* 108–126 (1969).
53. Balcan, D. *et al.* Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci.* **106**, 21484–21489 (2009).
54. Skellam, J. G. Random dispersal in theoretical populations. *Biometrika* **38**, 196–218 (1951).
55. Institut national de la statistique et des études économiques. Comparateur de territoire, <https://www.insee.fr/fr/statistiques/1405599?geo=DEP-75>.
56. Tokyo Metropolitan Government. Tokyo's history, geography, and population, <http://www.metro.tokyo.jp/ENGLISH/ABOUT/HISTORY/history03.htm>.
57. Pereira, R. H. M., Nadalin, V., Monasterio, L. & Albuquerque, P. H. Urban centrality: a simple index. *Geogr. analysis* **45**, 77–89 (2013).
58. Tsai, Y.-H. Quantifying urban form: compactness versus 'sprawl'. *Urban studies* **42**, 141–161 (2005).
59. Giuliano, G. & Small, K. A. Subcenters in the los angeles region. *Reg. science urban economics* **21**, 163–182 (1991).
60. Bertaud, A. & Malpezzi, S. The spatial distribution of population in 48 world cities: Implications for economies in transition. *Cent. for urban land economics research. Univ. Wis.* **32**, 54–55 (2003).
61. Louail, T. *et al.* From mobile phone data to the spatial structure of cities. *Sci. reports* **4**, 5276 (2014).
62. Boyle, P., Halfacree, K. & Robinson, V. Defining and measuring migration. *Explor. Contemp. Migr. Lond.* 34–56 (1998).
63. Shigesada, N. *et al.* Invasion and the range expansion of species: effects of long-distance dispersal. *Dispersal ecology* 350–373 (2002).
64. Afanasyev, V. *et al.* Revisiting lévy flight search patterns of wandering albatrosses, bumblebees and deer. *Nature* **449**, 6993Grimm (2007).
65. Petrovskii, S. & Morozov, A. Dispersal in a statistically structured population: fat tails revisited. *The Am. Nat.* **173**, 278–289 (2008).
66. McGrath Estate Agents. The mcgrath report 2020, https://issuu.com/mcgrathestateagents/docs/the_mcgrath_report_2020_-_mcgrath_estate_agents.
67. Sparke, E. *et al.* *Canberra 1954–1980* (The Australian National University, 1987).
68. Gibbney, H. J. *et al.* *Canberra 1913–1953* (The Australian National University, 1988).
69. Taylor, J. & Lea, J. P. Darwin. *Cities* **5**, 226–233 (1988).
70. Bureau of Meteorology. Tropical cyclone extremes, <http://www.bom.gov.au/cyclone/about/extremes.shtml>. Archived from the original on 21 July 2015. Retrieved 19 July 2015.
71. The SciPy community. SciPy: Open source scientific tools for Python (2001).

Acknowledgements

This work is supported by the University of Sydney's Postgraduate Research Scholarship SC0789, and the Australian Research Council Discovery Project DP170102927.

Author contributions

B.S. analyzed the data, all authors developed the model and prepared the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-65208-5>.

Correspondence and requests for materials should be addressed to B.S.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020