Decoding the city: multiscale spatial information of urban income

Luís M. A. Bettencourt[1,2,3*], Ivanna Rodriguez[1], Jordan T. Kemp[1,4], José Lobo[5]

[1*]Dept. of Ecology and Evolution, Dept. Sociology, University of Chicago, Chicago, 60637, IL, USA.
[2] Santa Fe Institute, Santa Fe, 87501, NM, USA.
[3] Vienna Complexity Science Hub, Vienna, 1030, Austria.
[4]Institute for New Economic Thinking at the Martin School, University of Oxford, Oxford , OX1 3UQ, UK.
[5]School of Sustainability, College of Global Futures, Arizona State University, Tempe, 85281, AZ, USA.

*Corresponding author(s). E-mail(s): bettencourt@uchicago.edu;

Abstract

Cities are characterized by the coexistence of general aggregate patterns, along with many local variations. This poses challenges for analyses of urban phenomena, which tend to be either too aggregated or too local, depending on the disciplinary approach. Here, we use methods from statistical learning theory to develop a general methodology for quantifying how much information is encoded in the spatial structure of cities at different scales. We illustrate the approach via the multiscale analysis of income distributions in over 900 US metropolitan areas. By treating the formation of diverse neighborhood structures as a process of spatial selection, we quantify the complexity of explanation needed to account for personal income heterogeneity observed across all US urban areas and each of their neighborhoods. We find that spatial selection is strongly dependent on income levels with richer and poorer households appearing spatially more segregated than middle-income groups. We also find that different neighborhoods present different degrees of income specificity and inequality, motivating analysis and theory beyond averages. Our findings emphasize the importance of multiscalar statistical methods that both coarse-grain and fine-grain data to bridge local to global theories of cities and other complex systems.

Keywords: Bayesian Statistics, Spatial Selection, Neighborhood Effects, Income

1 Introduction

Complex systems – such as cities, ecosystems, or organisms - are often recognizable by the coexistence of general structures along with many local variations [1, 2]. Such a description is deceptively simple though, because it glosses over the fact that local variations in a patch of forest or a city street represent not just random fluctuations [3, 4], as in physical systems, but a long history of serendipity and adaptation emphasized in biology and the social sciences [5, 6]. This distinction is central to any theoretical approach to cities: How to account for mechanisms of change and adaptation not only on large scales, but also locally in neighborhoods or small groups? What is chance and what is necessity in different places, at different times? How may local learning and adaptation lead to large-scale growth and development [7, 8]?

It is rather self-evident that the aggregate (or macro level) characteristics of an urban area should emerge from the interactions occurring within [9–11]. Recovering what these interactions are and identifying how they generate the macroscopic behavior is confused by the well-known "ecological fallacy" [12]. This issue results from inferences about the nature of individuals made on the basis of group averaged characteristics [2, 12]: Aggregating and averaging data invariably masks underlying heterogeneity and, in complex systems, may destroy relevant information.

The present paper develops a systematic approach to the analysis of how individual agent characteristics (such as income and other socioeconomic and demographic variables) form different patterns of organization at different scales. We use data from US metropolitan areas to motivate and illustrate our results, bridging large-scale patterns of income distribution at the metropolitan area scale (city) to those observed in small areas, such as block groups (neighborhoods), which are often quite different from the city at large.

The conceptual and empirical obstacles to building theories of complex systems across different scales are evident from the diverging approaches developed by different disciplines [3, 9, 10, 13]. For example, when modeling cities, physicists and economists tend to prefer the study of averaged behavior and adopt (homogeneous) representative agents [14], while other social scientists emphasize the specificity of places and people. Thus, from a statistical perspective, different approaches can be roughly divided into two camps. One emphasizes more aggregate data and modeling, while the other claims that such aggregation is (sometimes) a gross approximation to relevant phenomena and prefers local evidence and more attention to context and history. Here, our goal is not to adjudicate between these approaches but to demonstrate explicitly that both are necessary for the study of complex systems. In particular, we show that the complexity of explanation at different scales in cities can be quantified and pinpointed by appropriate methods of information theory.

Our approach contrasts with the common strategy in statistical physics –known as coarse-graining– of obtaining predictable macroscopic statistics from averaging over smaller and faster scales [15]. Coarse-graining is appropriate when variations on small and faster scales are random, resulting from microscopic disorder. Such approach then tells us that, in many known systems, most local details do not contribute to macroscopic behavior. This is the basis for the renormalization group in statistical physics,

which constitutes the essential tool to analyze phase-transitions in bulk materials and has resulted in powerful ideas of universality [13, 16].

It follows that proceeding in the direction of coarse-graining leads to information loss as local states are replaced by averages over larger scales [16], see Supplementary Text 1.1 for derivation. Because of this essential feature, renormalization group methods are not invertible. By contrast, proceeding in the opposite direction, from more averaged to less averaged systems – which we call "fine-graining"– requires the addition of information as new degrees of freedom on smaller scales must be specified. Such "fine-graining" methods have now been developed in statistical learning theory and inference [17–19], and in evolutionary biology in terms of selection [20, 21]. Following these developments, we see great opportunities for the use of emerging generative models from artificial intelligence to account for the detailed statistical structure of cities. Here, we lay some basic foundations for such future efforts by demonstrating and discussing some of the structures actually observed in US cities.

In pursuing an information-theoretic approach, we are also motivated by accounting in a new way for the complexity of human behavior in cities, and specifically the observed patterns of spatial sorting of households into neighborhoods by personal income [6, 22–24]. The differential sorting of households into neighborhoods, whether by income, ethnicity or any other characteristic, is a classic problem in sociology and economics [6, 22, 25, 26], approached originally by showing that seemingly innocuous local decisions can lead to extreme macroscopic patterns of segregation [26]. We will show, however, that observed patterns of economic sorting in US metropolitan areas are, generally, place and income group specific. This specifies how they entail more complex choices and therefore require different decision models to explain them in practice [27].

2 Results

To illustrate our methods and objectives, consider the pattern of household income in New York City neighborhoods (Figure 1A); see Figs. S1-S5 for details and other cities. We observe strong heterogeneity at different spatial scales, from adjacent neighborhoods (delineated by block groups, BK) with different average household incomes to larger patches of wealth and poverty, e.g. the Upper East Side (the richest part of Manhattan) or the Bronx (generally poor). Moreover, we observe different income distributions inside each neighborhood, Fig. 1B, so that poor neighborhoods contain people who are not poor (BK3), and rich neighborhoods many who are not rich (BK2). As is well known, this spatial heterogeneity is long-lived, persisting for decades or longer [6], through many economic cycles and substantial demographic turnover.

Interestingly, such rich and detailed patterns contrast with the simple normal distribution for (the logarithm of) household income across the entire city (metropolitan area), Fig. 1C. A log-normal process emerges from general multiplicative growth processes, where each contributing factor operates independently. In this case wealth appreciates at a fluctuating rate [28, 29]. This coarse-grained statistic is common to all US metropolitan areas: the distribution of income across all cities is well described

by a lognormal (except at the top tail, which is censored in this data), see Supplementary Text S1.2 and Figures S6-7. Moreover, the two parameters characterizing this distribution are themselves simple and general, see Supplementary Text 1.3 for background: The mean obeys a scaling relation [30, 31], well parameterized by a power-law $\langle Y(N,t))\rangle = Y_0(t)N^{\beta}(t)$, see Fig. S8, with general system-wide parameters $Y_0(t)$, β [2]; where the exponent $\beta > 1$, expresses urban superlinear (agglomeration) effects [30, 32], Fig. S8. Its value is predictable from urban scaling theory, which describes the city in terms of interdependent networks of people, organizations and infrastructure [2, 11]. The variance of the logarithmic income, see Fig. S9, is also a simple general number, associated with the cumulative volatility of incomes over time and setting the level of inequality within the urban area, as measured, for example, by the Gini coefficient or Theil's T index [29], Supplementary Text 1.4.

A statistical regularity thus emerges, Fig. 1C, at the city-wide scale as the result of averaging over a rich pattern of local neighborhood (and individual) variations, Fig. 1A-B. While this coarse-grained statistic reveals important aspects of urban socioeconomic dynamics [33, 34], we now focus on the structure of neighborhood variations using this macroscopic regularity as a reference. Specifically, we quantify the complexity of the pattern of variations at the neighborhood level by comparing income probability distributions at different levels of spatial aggregation, BK distributions in Fig.1B to Fig.1C. To do this explicitly, we write

$$p(y_\ell|n_j) = w_{\ell j}p(y_\ell) \tag{1}$$

where $p(y_\ell|n_j)$ is the distribution (normalized share) of income y, in discrete bins labeled by ℓ in neighborhood n_j (the different colorful BKs in Fig 1A). Here, $p(y_\ell)$ is the income distribution at a more aggregate level (Fig. 1C), which we take to be the entire city (metropolitan area) throughout this paper. Eq. 1 defines the weights, $w_{\ell j} \equiv p(y_\ell|n_j)/p(y_\ell)$ which transform one distribution into the other, see Fig. S10. With this definition, the average weights over income obey the normalization $\langle w_j\rangle = \sum_\ell w_{\ell j}p(y_\ell) = 1$ for all neighborhoods j, see Methods.

Eq. 1 is well known and can be readily recognized from two different perspectives. First, it is the haploid model of population genetics [5, 35], also known as the "replicator" equation in evolutionary game theory [36]. In that context, the two distributions are related across time (not space) and the weights $w_{\ell j}$ are the relative fitness of a trait ℓ, expressing its differential propagation to the next generation. The stronger the deviation of $w_{\ell j}$ away from the average (unity), the stronger the selection for allele ℓ. This corresponds to high fitness if $w_{\ell j} > 1$, and vice-versa if $w_{\ell j} < 1$. When $w_{\ell j} = 1$, the dynamics is neutral, and there is no selection. Selection refers to the process by which certain elements are favored over others based on specific features of the environment. The replicator equation allows the fitness function to incorporate the distribution of the population types rather than setting the fitness of a particular type constant, thus capturing the essence of selection [21]. This interpretation gives a mathematical correspondence between evolutionary dynamics (in time) and neighborhood sorting (in space).

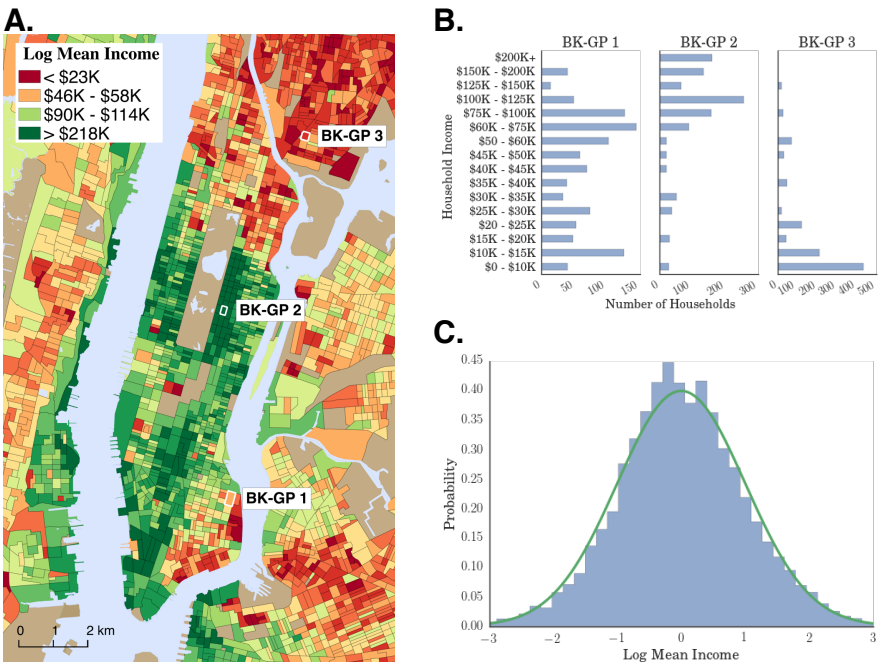

Fig. 1 The heterogeneity of neighborhoods in New York City. A. Average household income in New York City census block groups. B. Income distributions in selected neighborhood shown in Fig. 1A. C. The city-wide distribution of household income is well described by a lognormal distribution (green line), which we show in SOM is a very good general model for the household income distribution in all US MSAs. The mean and variance of this distribution obey simple scaling relations, see Supplementary Materials **S2.2**. Data was compiled by the US Census 5-year American Community Survey (2010 release) comprising of over 200,000 block groups nationwide and about 14,000 in the New York Metropolitan Statistical Area (MSA).

Second, Eq. 1 is a form of Bayes' theorem [17], which leads to the interpretation of $w_{\ell j}$ in terms of probability ratios, specifically

$$p(y_\ell|n_j) = \frac{p(n_j|y_\ell)}{p(n_j)}p(y_\ell) \quad \rightarrow \quad w_{\ell j} = \frac{p(n_j|y_\ell)}{p(n_j)} = \frac{p(y_\ell|n_j)}{p(y_\ell)} = \frac{p(y_\ell, n_j)}{p(y_\ell)p(n_j)}. \tag{2}$$

Here, $p(n_j|y_\ell)$ is the probability for a person in the city to reside in neighborhood n_j, given that they have income y_ℓ. The probability, $p(n_j)$, is the (income independent) probability to live in neighborhood n_j, estimated as the ratio of its population to that of the city, see Methods.

This second perspective leads to another powerful correspondence between probability theory, inference, and neighborhood structure. In this context, $\log w_{\ell j}$, in Eq. 2 is the (non-averaged) mutual information [17] between neighborhood j and the distribution of income y. (Mutual information measures how much knowing one random variable reduces uncertainty about another, quantifying the amount of shared information between them.) To see this more explicitly consider the average of $\log w_{\ell j}$ over

income groups

$$\langle \log w_j \rangle = \sum_l p(y_\ell|n_j) \log \frac{p(y_\ell|n_j)}{p(y_\ell)} = D[p(y|n_j||p(y)]. \quad (3)$$

Here, $D[...]$ is the (Kullback-Leibler) divergence between the distributions of income city-wide and in neighborhood j. The Kullback-Leibler divergence is a core concept in information theory that quantifies how much one probability distribution differs from another, expected distribution—effectively measuring the information lost when the latter is used to approximate the former [37]. For each neighborhood j, it represents the amount of information required to describe its specific income distribution, assuming we begin with knowledge of the city-wide income pattern. Atypical neighborhoods, with income distributions very different from the city as a whole, require a longer explanation (more information). Neighborhoods that already reflect the city-wide pattern require no further description. In other words, atypical neighborhoods necessitate the invocation of local neighborhood effects, in addition to the city-wide distribution of traits common to all places, to explain their income distributions. The term $\langle \log w_j \rangle$ expresses the strength of neighborhood effects in each j relative to city-wide dynamics, measured in units of information.

Fig. 2A shows the strength of income neighborhood effects for each block-group in New York City, measured by $\langle \log w_j \rangle$, see Figs. S11-S15 for details and other cities. We observe a very mixed pattern of local selection with many neighborhoods reflecting the distribution of income for the city as a whole (dark gray). However, others display a strong local flavor (red). We verified that the magnitude of the observed differences could not be the result of a purely random process of drawing individuals at random from the metropolitan income distribution into each place, Fig. S16.

Comparing Figs. 1A and 2A suggests that the most atypical neighborhoods have both the highest and the lowest average household incomes. It turns out that this is a general pattern of selection across all US metropolitan areas that we can quantify systematically via the average of $\log w_{\ell j}$ over neighborhoods j,

$$\langle \log w_\ell \rangle = \sum_j p(n_j|y_\ell) \log \frac{p(n_j|y_\ell)}{p(n_j)} = D[p(n|y_\ell)||p(n)]. \quad (4)$$

This quantity is the average information necessary to explain the distribution of specific income groups y_ℓ across the city, given that we know its neighborhood structure. In the absence of neighborhood effects (i.e., of spatial sorting), this quantity is zero, meaning that each level of income is distributed at random over space. Thus, the magnitude of Eq. 4 quantifies the differential average strength of neighborhood effects for different income levels in each city. Fig. 2B shows that the neighborhood effects are strongest for the highest income group, followed by the lowest. Middle-incomes are observed to be spatially the most mixed and thus less determined by specific neighborhoods. This is an interesting finding because it shows that different income groups

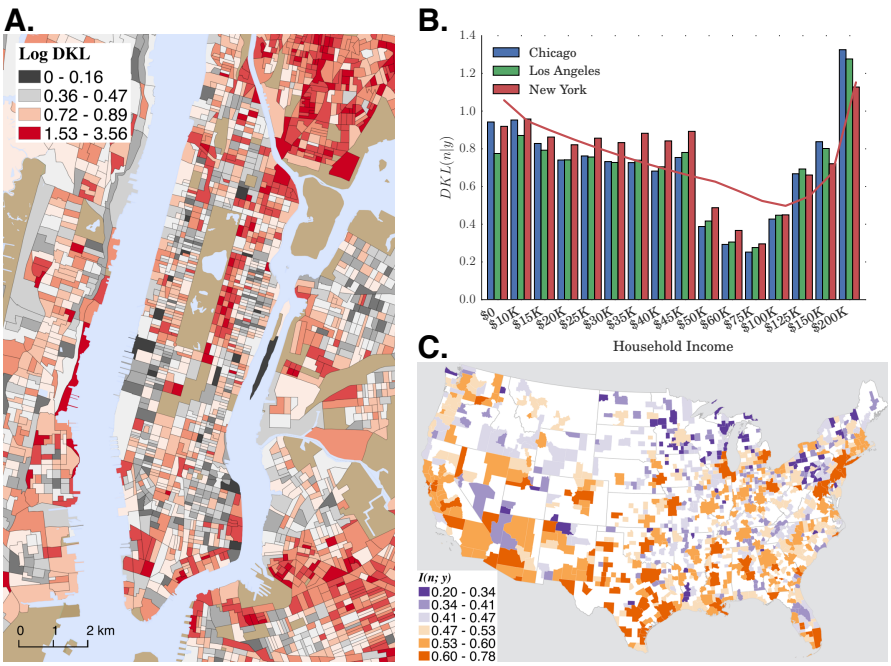

Fig. 2 Spatial Selection Measured in Units of Information. A. The information, $\langle \log w_j \rangle$, necessary to explain the household income distribution of different neighborhoods, given the city-wide income statistics, Fig. 1C. The entropy of the income distribution for the New York Metropolitan Area is $H(y) = 3.89$ bits. Thus neighborhoods in darker gray require very little additional information, while those in red may demand local explanations that are as comprehensive as the city-wide pattern of income itself. Comparing to Fig. 1A, note that it is both the poorest and richest neighborhoods that tend to require more information (red). B. The average intensity of local selection by income, measured by $\langle \log w_l \rangle$, for several large US MSAs. The richest income brackets experience the strongest spatial selection, followed by the lowest incomes. The variation with income can be reasonably fitted by a quadratic form (red line, for New York MSA) with $\langle \log w_l \rangle = 1.058 - 1.16710^{-5} y + 6.074 x 10^{-11} y^2$. C. The average strength of neighborhood effects across urban areas in the US measured by the mutual information $I(y;n)$ (942 metropolitan and micropolitan areas). Orange denotes stronger average neighborhood selection, where the distribution of income in each of the city's neighborhoods is less like that of the city as a whole, and vice versa (purple). Cities with low $I(y;n) \to 0$ have less distinguishable neighborhood structure by income.

exercise different kinds of choices – by preference and necessity - in terms of residential location. This also demonstrates that any realistic model of residential choice in US cities needs to be an explicit function of income levels.

These two effects are summarized in turn by a single quantity that captures the overall strength of neighborhood effects for each city in units of information, Fig. 2C. This is the total (mutual) information, $I(y;n) = \langle \log w \rangle$, between neighborhood structure and income, given as the average of the previous quantities over the remaining variable,

$$\langle \log w \rangle \quad = \sum_j p(n_j) D[p(y|n_j)||p(y)] = \sum_\ell p(y_\ell) D[p(n|y_\ell)||p(n)] \tag{5}$$

$$= \sum_{\ell,j} p(y_\ell, n_j) \log w_{\ell j} = I(y;n).$$

If every neighborhood were a microcosm of the city as a whole, then all income groups would be spatially well-mixed and there would be no (income) neighborhood effects, leading to $I(y;n) = 0$. Conversely, in cities where every neighborhood has its own unique flavor, not at all like the distribution of traits across the city, there is strong sorting of incomes by neighborhood and $I(y;n)$ will be large. How large depends on the relative amount of information needed to describe the system at the local level, Fig. 1B, versus as a whole, Fig. 1C. The mutual information $I(y;n)$ gives a measure of how well a coarse-grained pattern describes a complex system observed at a more disaggregated level. In other words, $I(y;n)$ quantifies the average complexity of any theory of local neighborhood effects versus a theory of the same quantity at the metropolitan level.

The top and bottom ranked metropolitan areas in the US by the magnitude of $I(y,n)$ are shown in Tables S1-S4. We see that Dallas TX, followed by New York City and New Orleans, LA show the highest $I(y;n)$ in 2010 and that many cities in Texas show in general strong income segregation by neighborhood. This is particularly interesting because these cities are currently among the fastest growing in the nation so that some of the observed income segregation is the result of recent residential choices. Smaller cities, especially in parts of the Midwest (e.g. Wisconsin) but also in other states, show the lowest neighborhood segregation by income.

Figure 3A shows how the strength of neighborhood effects measured by $I(y;n)$ changes over time, for select larger cities. Note that this analysis is done for three broad temporal periods – 2010, 2015, and 2020 – because the data is collected on a 5-year rolling basis. We observe very similar results for 2010 and 2015, but a sharp increase in neighborhood sorting effects in 2020. This effect seems to be more pronounced in some larger cities and is associated with more dispersed observations across all cities, as shown in Fig. 3B. While we do not have a simple explanation for these observations, the latter part of the 5-year American Community Survey coincided with the COVID-19 pandemic. On the one hand, the pandemic is known to have caused significant non-response bias in the 2020 data in this survey [38]. On the other hand, there were significant (temporary) population dislocations, with many (more affluent) households leaving central areas of larger cities and creating potentially strong selection effects [39–41]. It will be interesting to continue to monitor this situation to establish the reason and potential reversal of these effects over time.

3 Methods

3.1 Data Sources

Geo-referenced data at the household level (income and population) for the United States is reported at the *Census Block Group* level (BK) by the 5-year American Community Survey. Block groups are statistical subdivisions of Census Tracts, which in turn are the basic data collection units for the population census. The boundaries of block groups are generally set so that they contain between 600 and 3,000

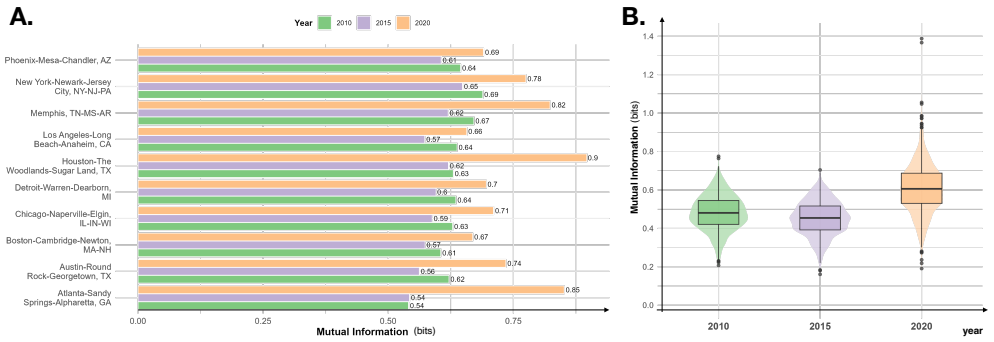

Fig. 3 Mutual information between income and neighborhoods over time. A. For select large Metropolitan Areas. B. Mutual information across cities over time. The middle line shows the median across cities. Boxes show the inter-quartile range (between the first and third quartiles). Some outliers are shown as points. Shaded areas show the shape of the distribution.

people, with a typical size of about 1,500 (or 500 households). Block Groups are spatially contiguous and tile the entire country. Data are aggregated into urban areas defined as Core-Based Statistical Areas (CBSA), which include Micropolitan Statistical Areas and Metropolitan Statistical Areas. Metropolitan areas contain an urban core of 50,000 or more people, while micropolitan areas center around an urban core with 10,000 to 49,999 residents, both defined by adjacent counties with strong economic and commuting ties.(An urban core is a densely settled area.) For simplicity we refer to micropolitan and metropolitan areas together as *Metropolitan Areas*: there are 942 such areas currently in the USA. Counties are the primary legal divisions of States in the U.S., many of which are functioning governmental units whose powers and functions vary from state to state. Counties differ greatly in their areal expansion and populations size.

3.2 Units of Analysis: Neighborhood Definitions

In the main text, we use the term *neighborhood* to refer to Census Block Groups, as shown in Figure 1A. Block groups provide an exhaustive tiling of the entire national territory of the United States and its population. In denser areas, Block Groups correspond to smaller land areas, as can be clearly seen in the maps of Fig.1A. Adopting block groups as proxies for neighborhoods is convenient because they are consistently defined by the US Census Bureau (and similarly by other national census around the world) and provide a universal standard for the study of small area statistics across an entire nation. For these reasons, they are the most common proxies for social units at this scale (neighborhoods) in the USA.

A *neighborhood* is commonly defined as a geographically localized community where residents share a collective identity, social interactions, and access to local resources. To empirically operationalize this socio-spatial concept, however, a standardized spatial unit is required, and in this study, we follow the common convention of using Census Block Groups as a proxy for neighborhoods (as depicted in Figure

1A). This approach is advantageous because the U.S. Census Bureau intentionally designs block groups to be relatively homogeneous in population and housing, and their exhaustive, standardized coverage of the entire nation enables robust, replicable analysis, making them the default choice for large-scale quantitative research. Nonetheless, this choice has a significant drawback: these administrative boundaries are externally imposed and may not align with residents' perceived community lines or social geographies, potentially bisecting cohesive groups or artificially combining disparate ones. This fundamental tension between administrative convenience and lived experience is why sociologists have long debated the validity of block groups, arguing they can obscure nuanced social dynamics and prompting many local studies to adopt different, more socially meaningful units of analysis (see for example [42]).

Our aim here is to demonstrate effects of spatial selection at any given scale. A systematic study of the strength of spatial selection at different scales using different delineations of neighborhoods is beyond the scope of the present manuscript and will be presented elsewhere.

3.3 Data limitations

The American Community Survey (ACS) and the Decennial Census collect household data in small spatial units that allow us to characterize patterns of spatial selection in neighborhoods (i.e., Census Blocks). The ACS is a statistical survey conducted by the US Census Bureau, sent to approximately 250,000 addresses monthly (or about 3 million per year). Unlike the population census (which is strictly a population count), the ACS collects socioeconomic information (for example, on household income). The data are collected primarily by mail, with follow-ups by telephone and personal visits. ACS data are used to make yearly estimates for counties which are then aggregated to provide estimates for States and metropolitan areas.[1] ACS data has an important reporting limitation when it comes to the upper tail of the income distribution: the number of households is listed only for a data bin set by a minimum value ($> \$200$k per household in 2010).

It has been often shown empirically [43] that, at higher levels of spatial aggregation, the upper tail income distribution deviates from the lognormal pattern reported in Fig. 1C. Such statistics do, in fact, often follow a Pareto (power-law) distribution for the top richest fraction of 1% [43]. Consideration of a finer distribution in this regime is likely to produce even higher atypical values of information for neighborhoods that concentrate such high incomes. In this sense, even though many richer neighborhoods appear the most atypical from the point of view of their income distribution relative to the city at large, Fig. 2B, it is likely that this effect is underestimated as a result of the way data for these incomes are reported.

3.4 Practical Estimation of Probabilities

Here, we provide an explicit version of the probability distributions introduced in the main text and the procedure by which they are estimated from discretely binned data.

[1] For detailed information on the American Community Survey go to www.census.gov/acs/.

Let N be the the total number of households in a given city, or the size of that city, for short. Let N_j be the number of households in neighborhood j, across all income levels. Then $n_{j,\ell}$ is the number of households in neighborhood j, with income (in the interval denoted by) ℓ. N_ℓ is, correspondingly, the total number of households in the city with income in the interval indexed by ℓ. These quantities obey several simple sum rules:

$$\sum_j N_j = N, \quad \sum_\ell N_\ell = N, \qquad \sum_j n_{j,\ell} = N_\ell, \quad \sum_\ell n_{j,\ell} = N_j. \tag{6}$$

Having defined these quantities, which are the ones typically reported by the U.S. Census Bureau, we can provide simple frequency estimators for the several probability densities introduced in the main manuscript. The simplest is $p(n_j)$, the probability of living in a specific neighborhood, which is $p(n_j) = \frac{N_j}{N}$. Analogously, the probability of belonging to a given income level, ℓ, is $p(y_\ell) = \frac{N_\ell}{N}$. The conditional distribution for being in a given neighborhood j given income ℓ is $p(n_j|y_\ell) = \frac{n_{j,\ell}}{N_j}$. From this and Bayes' relation it follows that

$$p(y_\ell|n_j) = \frac{p(n_j|y_\ell)}{p(n_j)} p(y_\ell) = \frac{n_{j,\ell}}{N_j}. \tag{7}$$

The weights $w_{j,\ell}$ are given by

$$w_{j,\ell} = N \frac{n_{j,\ell}}{N_\ell N_j}. \tag{8}$$

Finally, we can check that the properties of the conditional probabilities hold, under these definitions,

$$\sum_j p(n_j|y_\ell) = \sum_j \frac{n_{j,\ell}}{N_\ell} = \frac{1}{N_\ell} N_\ell = 1. \tag{9}$$

$$\sum_\ell p(n_j|y_\ell) N_\ell = \sum_\ell \frac{n_{j,\ell}}{N_\ell} N_\ell = \sum_\ell n_{j,\ell} = N_j. \tag{10}$$

$$\sum_\ell p(y_\ell|n_j) = \sum_\ell \frac{n_{j,\ell}}{N_j} = \frac{1}{N_j} N_j = 1. \tag{11}$$

$$\sum_j p(y_\ell|n_j) N_j = \sum_j \frac{n_{j,\ell}}{N_j} N_j = \sum_j n_{j,\ell} = N_\ell. \tag{12}$$

4 Discussion

We have shown how intricate local patterns of population traits can develop from broader, more general distributions in larger populations through processes of spatial selection. We also demonstrated that the type and strength of these patterns are most effectively measured in terms of information—a fundamental quantity that provides a unifying language for analyzing complex systems. Cities display simple aggregate

regularities alongside rich local variation (often described as "neighborhoods"). Our central contribution is to show that spatial sorting of households across neighborhoods can be represented, measured, and interpreted as a selection process whose intensity is meaningfully quantified in units of information.

Selection is a general process by which populations and individuals adapt to their environment by acquiring and processing information and revealing their choices [20, 21, 44]. Although the word "selection" may evoke biological evolution, here it formalizes a general mechanism by which systems acquire information by differentially amplifying some types over others. As Price noted more than 50 years ago [45], selection is broadly applicable and can be described by the same mathematical formalism. When applied to spatial sorting, this approach allows us to study how local variations can arise within the context of widespread statistical regularities.

In the context of cities, the "types" are income groups, the "environment" is the configuration of housing, amenities, institutions, prices, and social networks, and the amplification arises from households' choices and constraints interacting with supply, policy, discrimination, and path dependence. The replicator–Bayes equivalence in Eqs. 1–2 makes this explicit: residential choice behaves as inference, where the neighborhoods signal rewards and restrictions, the households update beliefs and options, and the urban system learns by reweighting the joint distribution (y, n) towards greater compatibility.

The quantification of processes of selection in terms of information is still relatively new [20, 21, 35]. We hope that the results described here provide new additional perspectives into this fundamental identification in respect to the observed multiscale distribution of population and associated features over space in cities. To this end, we have shown how to account for information embedded in the spatial structure of cities at different scales from metropolitan areas to local neighborhoods, thus accounting for the complexity of an overall pattern in terms of both local and global mechanisms. In this way, we hope to bridge two frequently opposing perspectives, by showing how coarse-grained "universality" can co-exist with local choice and specificity. This is particularly poignant for cities, where diversity at the neighborhood levels coexists with relatively simple regularities at the level of functional cities and urban systems [2, 6, 10]. The application of these methods to other locally heterogeneous systems, such as ecosystems or neural networks, is straightforward but requires datasets of comparable scope and quality.

Models of residential choice have recently been developed representing salient and empirically supported household decisions, characterized by more realistic local contexts, personal traits, and continuous levels of preference [23, 27, 46]. Estimating these personal and contextual characteristics from empirical data requires a systematic methodology such as the one introduced here, which at once measures the complexity (i.e. the amount of information) necessary at different scales. The recursive property of informational quantities tells us how much of the explanation for an overall pattern may be macroscopic or microscopic by helping us keep track of the information content of models at different scales, Supplementary Text 1.4. Our results show specifically that the aggregate distribution of income for metropolitan New York City is a poor model for most of the city's neighborhoods, especially its richest and poorest places.

A limitation of the present study is related to the quality of data over time and also to other quantities besides income. As we have seen for the US, this data is created via a rolling survey, which is sensitive to non-response biases in times of fast change, such during the recent COVID-19 pandemic. In this respect, studies of neighborhood effects related to income using register-based data in nations such as the Netherlands [47], Sweden[48, 49], Norway [50, 51], or Finland [52, 53] present much richer empirical opportunities, also to analyze changes over time and throughout people's life course towards general theory.

A more systematic understanding of spatial population sorting by personal income and other household characteristics remains at the root of some of the most challenging problems for urban science and policy, including the causes and consequences of economic inequality [54–56], ethnic and racial segregation, disparate access to opportunity [56–58] and spatially concentrated (dis)advantage [51, 59, 60], including issues of crime and violence [6, 24, 59, 61], Supplementary Text 1.5. Extensions of present models and analytical approaches to more urban systems (nations) and several other demographic dimensions (income, race, education, gender, etc) remain necessary to create better theory of human development and change, and to make urban policy and practice more effective in the face of radically different challenges faced by specific individuals in different places.

Supplementary information. Includes supplementary text, sixteen figures and four tables.

Acknowledgements. We thank Christa Brelsford, Elisabeth Bruch, and Laura Fürsich for discussions. This research was partially supported by the Mansueto Institute for Urban Innovation at the University of Chicago, and by the Arizona State University/Santa Fe Institute Center for Biosocial Complex Systems.

Declarations

- Funding: This work was partially supported by the Mansueto Institute at the University of Chicago.
- Competing interests: The authors declare no competing interests.
- Data, Materials and Code availability; Data, code and materials are available at https://github.com/mansueto-institute/dkl-metric
- Author contribution: LB and JL conceptualized the study and methods, IR performed data curation and computation, LB, JL, JK wrote the paper.

References

[1] Goldenfeld, N., Kadanoff, L.P.: Simple Lessons from Complexity. Science **284**(5411), 87–89 (1999) https://doi.org/10.1126/science.284.5411.87 . Accessed 2015-06-14

[2] Bettencourt, L.M.A.: Introduction to Urban Science. The MIT Press, Cambridge, MA (2021). https://mitpress.mit.edu/9780262046003/introduction-to-urban-science/

[3] Bettencourt, L.M.A., 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**(11), 13541 (2010) https://doi.org/10.1371/journal.pone.0013541

[4] Bettencourt, L.M.A.: Towards a statistical mechanics of cities. Comptes Rendus Physique (2019) https://doi.org/10.1016/j.crhy.2019.05.007 . Accessed 2019-07-18

[5] Crow, J.F., Kimura, M.: An Introduction to Population Genetics Theory. Harper & Row, New York (1970)

[6] Sampson, R.J.: Great American City: Chicago and the Enduring Neighborhood Effect. University of Chicago Press, Chicago, IL (2012). https://doi.org/10.7208/chicago/9780226733883.001.0001 . https://doi.org/10.7208/chicago/9780226733883.001.0001

[7] Jones, C.I., Romer, P.M.: The New Kaldor Facts: Ideas, Institutions, Population, and Human Capital. American Economic Journal: Macroeconomics **2**(1), 224–245 (2010) https://doi.org/10.1257/mac.2.1.224 . Accessed 2024-09-10

[8] Kemp, J.T., Kline, A.G., Bettencourt, L.M.A.: Information synergy maximizes the growth rate of heterogeneous groups. PNAS Nexus **3**(2), 072 (2024) https://doi.org/10.1093/pnasnexus/pgae072

[9] Jacobs, J.: The Death and Life of Great American Cities, 50th anniversary ed., 2011 modern library ed edn. Modern Library, New York (2011). OCLC: ocn748543003

[10] Batty, M.: The size, scale, and shape of cities. Science **319**(5864), 769–771 (2008) https://doi.org/10.1126/science.1151419

[11] Bettencourt, L.M.A.: The origins of scaling in cities. Science **340**(6139), 1438–1441 (2013) https://doi.org/10.1126/science.1235823

[12] Robinson, W.S.: Ecological correlations and the behavior of individuals. American Sociological Review **15**(3), 351–357 (1950). Accessed 2025-09-16

[13] Anderson, P.W.: More Is Different. Science **177**(4047), 393–396 (1972) https://doi.org/10.1126/science.177.4047.393 . Accessed 2018-02-05

[14] Axtell, R.L., Farmer, J.D.: Agent-based modeling in economics and finance: Past, present, and future. J. Econ. Lit. **63**(1), 197–287 (2025) https://doi.org/10.1257/jel.20221601

[15] Goldenfeld, N.: Lectures on Phase Transitions and the Renormalization Group. Frontiers in physics, vol. v. 85. Addison-Wesley, Advanced Book Program, Reading, Mass (1992)

[16] Kadanoff, L.P.: Statistical Physics: Statics, Dynamics and Renormalization. World Scientific, Singapore ; River Edge, N.J (2000)

[17] MacKay, D.J.C.: Information Theory, Inference, and Learning Algorithms. Cambridge University Press, Cambridge, UK ; New York (2003)

[18] Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: European Conference on Computer Vision (ECCV), pp. 184–199 (2014). https://doi.org/10.1007/978-3-319-10593-2_13

[19] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., Shi, W.: Photo-realistic single image super-resolution using a generative adversarial network. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 105–114 (2017). https://doi.org/10.1109/CVPR.2017.19

[20] Donaldson-Matasci, M.C., Bergstrom, C.T., Lachmann, M.: The fitness value of information. Oikos **119**(2), 219–230 (2010) https://doi.org/10.1111/j.1600-0706.2009.17781.x

[21] Bettencourt, L.M.A., Grandison, B.J., Kemp, J.T.: Redefining Fitness: Evolution as a Dynamic Learning Process (2025). https://doi.org/10.48550/arXiv.2503.09057 . https://arxiv.org/abs/2503.09057

[22] Wilson, W.J.: The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy. University of Chicago Press, Chicago, IL (1987)

[23] Bruch, E.E.: How population structure shapes neighborhood segregation. American Journal of Sociology **119**(5), 1221–1278 (2014) https://doi.org/10.1086/675411

[24] Intrator, J., Tannen, J., Massey, D.S.: Segregation by race and income in the united states 1970–2010. Social Science Research **60**, 45–60 (2016) https://doi.org/10.1016/j.ssresearch.2016.08.003

[25] Park, R.E., Burgess, E.W., McKenzie, R.D., Janowitz, M.: The City: Suggestions for Investigation of Human Behavior in the Urban Environment, Nachdr. edn. The heritage of sociology. Univ. of Chicago Press, Chicago (2010). OCLC: 837592928

[26] Schelling, T.C.: Dynamic models of segregation. The Journal of Mathematical Sociology **1**(2), 143–186 (1971) https://doi.org/10.1080/0022250X.1971.9989794 . Accessed 2019-08-28

[27] Bruch, E.E., Mare, R.D.: Neighborhood choice and neighborhood change. American Journal of Sociology **112**(3), 667–709 (2006) https://doi.org/10.1086/507856

[28] Gabaix, X.: Zipf's Law for Cities: An Explanation. The Quarterly Journal of Economics **114**(3), 739–767 (1999) https://doi.org/10.1162/003355399556133 . Accessed 2015-08-24

[29] Kemp, J.T., Bettencourt, L.M.A.: Statistical dynamics of wealth inequality in stochastic models of growth. Physica A: Statistical Mechanics and its Applications **607**, 128180 (2022) https://doi.org/10.1016/j.physa.2022.128180 . Accessed 2025-08-18

[30] Bettencourt, L.M.A., Lobo, J., Helbing, D., Kühnert, C., West, G.B.: Growth, innovation, scaling, and the pace of life in cities. Proc. Natl. Acad. Sci. U.S.A. **104**(17), 7301–7306 (2007) https://doi.org/10.1073/pnas.0610172104

[31] Gomez-Lievano, A., Youn, H., Bettencourt, L.M.A.: The statistics of urban scaling and their connection to zipf's law. PLoS ONE **7**(7), 40393 (2012) https://doi.org/10.1371/journal.pone.0040393

[32] Bettencourt, L.M.A., Lobo, J., Youn, H.: The hypothesis of urban scaling: formalization, implications and challenges. arXiv:1301.5919 [nlin, physics:physics] (2013). arXiv: 1301.5919. Accessed 2015-08-24

[33] Schlapfer, M., Bettencourt, L.M.A., Grauwin, S., Raschke, M., Claxton, R., Smoreda, Z., West, G.B., Ratti, C.: The scaling of human interactions with city size. Journal of The Royal Society Interface **11**(98), 20130789–20130789 (2014) https://doi.org/10.1098/rsif.2013.0789 . Accessed 2015-03-23

[34] Bettencourt, L.M.A.: Urban growth and the emergent statistics of cities. Science Advances **6**(34), 8812 (2020) https://doi.org/10.1126/sciadv.aat8812 . Accessed 2021-08-26

[35] Frank, S.A.: Natural selection. I. Variable environments and uncertain returns on investment. Journal of Evolutionary Biology **24**(11), 2299–2309 (2011) https://doi.org/10.1111/j.1420-9101.2011.02378.x

[36] Page, K.M., Nowak, M.A.: Unifying evolutionary dynamics. J. Theor. Biol. **219**(1), 93–98 (2002) https://doi.org/10.1006/jtbi.2002.3112

[37] Cover, T.M., Thomas, J.A.: Elements of Information Theory, 2nd edn. Wiley Series in Telecommunications and Signal Processing. Wiley-Interscience, Hoboken, NJ (2006). https://doi.org/10.1002/047174882X . https://doi.org/10.1002/047174882X

[38] Jon, R., Eggleston, J., Bee, A., Klee, M., Mendez-Smith, B.: Addressing nonresponse bias in the american community survey during the pandemic using administrative data. Technical report, U.S. Census Bureau (2021). https://www.census.gov/library/working-papers/2021/acs/2021_Rothbaum_01.html

[39] Coven, J., Gupta, A., Yao, I.: Urban flight seeded the covid-19 pandemic across the united states. JUE Insight **133**(103489) (2023) https://doi.org/10.1016/j.jue.2022.103489

[40] Foster, T.B., Fiorio, L., Ellis, M.: Internal migration in the u.s. during the covid-19 pandemic. Technical Report CES Working Paper 24-50, U.S. Census Bureau, Center for Economic Studies (2024). https://www2.census.gov/library/working-papers/2024/adrm/ces/CES-WP-24-50.pdf

[41] Rebhun, U., Brown, D.L.: The covid-19 pandemic in the united states: Who moved, where, and why? Population Research and Policy Review **Year 2025**, 44–16 (2025) https://doi.org/10.1007/s11113-024-09928-w

[42] Hipp, J.R.: Block, tract, and levels of aggregation: Neighborhood structure and crime and disorder as a case in point. Am. Sociol. Rev. **72**(5), 659–680 (2007) https://doi.org/10.1177/000312240707200501

[43] Montroll, E.W., Shlesinger, M.F.: On 1/¡i¿f¡/i¿ noise and other distributions with long tails. Proceedings of the National Academy of Sciences **79**(10), 3380–3383 (1982) https://doi.org/10.1073/pnas.79.10.3380 https://www.pnas.org/doi/pdf/10.1073/pnas.79.10.3380

[44] Hilbert, M.: The More You Know, the More You Can Grow: An Information Theoretic Approach to Growth in the Information Age. Entropy **19**(2), 82 (2017) https://doi.org/10.3390/e19020082 . Accessed 2017-06-19

[45] Price, G.R.: Selection and Covariance. Nature **227**(5257), 520–521 (1970) https://doi.org/10.1038/227520a0 . Accessed 2017-05-08

[46] Reardon, S.F., Bischoff, K.: Income inequality and income segregation. Am. J. Sociol. **116**(4), 1092–1153 (2011) https://doi.org/10.1086/657114

[47] Aline, T.A., Ham, M., Tammaru, T., Musterd, S.: Neighbourhood effects on educational attainment: What matters more, poverty or affluence? PLOS ONE **18**(3), 0281928 (2023) https://doi.org/10.1371/journal.pone.0281928

[48] Galster, G., Andersson, R., Musterd, S.: Who is affected by neighbourhood income mix? gender, age, family, employment and income differences. Urban Studies **47**(14), 2915–2944 (2010) https://doi.org/10.1177/0042098009360233 https://doi.org/10.1177/0042098009360233

[49] Hedman, L., Manley, D., Ham, M., Östh, J.: Cumulative exposure to disadvantage and the intergenerational transmission of neighbourhood effects. Journal of Economic Geography **15**(1), 195–215 (2015) https://doi.org/10.1093/jeg/lbt042

[50] Brattbakk, I., Wessel, T.: Long-term neighbourhood effects on education, income and employment among adolescents in oslo. Urban Studies **50**(2), 391–406 (2013)

https://doi.org/10.1177/0042098012448548

[51] Borgen, N.T., Hermansen, A.S., *et al.*: How neighbourhood effects vary by achievement level. European Sociological Review **41**(2), 215–233 (2025) https://doi.org/10.1093/esr/jcad116

[52] Tarkiainen, L., Martikainen, P., *et al.*: Long-term neighbourhood inequalities in cause-specific mortality and hospitalisation: multilevel analyses among individuals nested in finnish post-code areas, 1991-2018. SSM Population Health **23**, 101323 (2022) https://doi.org/10.1016/j.ssmph.2022.101323

[53] Ristikari, T., *et al.*: The effect of cumulative childhood exposure to neighbourhood socioeconomic disadvantage on school performance: a register-based study on neighbourhoods, schools and siblings. European Sociological Review **40**(3), 403–420 (2024) https://doi.org/10.1093/esr/jcad020

[54] Krivo, L.J., Washington, H.M., Peterson, R.D., Browning, C.R., Calder, C.A., Kwan, M.-P., Lee, J.-Y., Matthews, S.A.: Social isolation of disadvantage and advantage: The reproduction of inequality in urban space. Soc. Forces **92**(1), 141–164 (2013) https://doi.org/10.1093/sf/sot043

[55] Chen, W.-H., Myles, J., Picot, G.: Why have poorer neighbourhoods stagnated economically while the richer have flourished? neighbourhood income inequality in canadian cities. Urban Stud. **49**(4), 877–896 (2012) https://doi.org/10.1177/0042098011408142

[56] Lens, M.C.: Measuring the geography of opportunity. Progress in Human Geography **41**(1), 3–25 (2017) https://doi.org/10.1177/0309132515618104 . Accessed 2017-05-07

[57] Chetty, R., Grusky, D., Hell, M., Hendren, N., Manduca, R., Narang, J.: The fading american dream: Trends in absolute income mobility since 1940. Science **356**(6336), 398–406 (2017) https://doi.org/10.1126/science.aal4617

[58] Chetty, R., Jackson, M.O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R.B., Gong, S., Gonzalez, F., Grondin, A., Jacob, M., Johnston, D., Koenen, M., Laguna-Muggenburg, E., Mudekereza, F., Rutter, T., Thor, N., Townsend, W., Zhang, R., Bailey, M., Barberá, P., Bhole, M., Wernerfelt, N.: Social capital II: determinants of economic connectedness. Nature **608**(7921), 122–134 (2022) https://doi.org/10.1038/s41586-022-04997-3 . Accessed 2025-08-17

[59] Wilson, W.J.: Studying inner-city social dislocations: The challenge of public agenda research: 1990 presidential address. American Sociological Review **56**(1), 1–14 (1991) https://doi.org/10.2307/2095675

[60] Elliott, D.S., Wilson, W.J., Huizinga, D., Sampson, R.J., Elliott, A., Rankin, B.: The Effects of Neighborhood Disadvantage on Adolescent Development. Journal

of Research in Crime and Delinquency **33**(4), 389–426 (1996) https://doi.org/10.1177/0022427896033004002 . Accessed 2016-02-16

[61] Besbris, M., Faber, J.W., Rich, P., Sharkey, P.: Effect of neighborhood stigma on economic transactions. Proc. Natl. Acad. Sci. U.S.A. **112**(16), 4994–4998 (2015) https://doi.org/10.1073/pnas.1414139112