Distribution of Neighbourhood Size in Cities

Anand Sahasranaman and Henrik Jensen

July 11, 2021

Distribution of neighborhood size in cities

Anand Sahasranaman[1,2,*] and Henrik Jeldtoft Jensen[2,3#]

[1]Division of Science, Division of Humanities and Social Science, Krea University, Andhra Pradesh 517646, India.

[2]Centre for Complexity Science, Dept of Mathematics, Imperial College London, London SW72AZ, UK.

[3]Institute of Innovative Research, Tokyo Institute of Technology, 4259, Nagatsuta-cho, Yokohama 226-8502, Japan

[*] Corresponding Author. Email: anand.sahasranaman@krea.edu.in

[#] Email: h.jensen@imperial.ac.uk

Abstract:

While the distribution of city sizes in a nation is a well explored question, there is little work on understanding the distribution of neighbourhood sizes within cities. We seek to explore the size distributions of neighbourhoods in cities and propose a candidate explanation for emergent distribution patterns. We also explore the consistency of our empirical findings on neighbourhood size distributions with the Zipfian distribution of city sizes. We use neighbourhood level data from 12 global metropolises to statistically characterize neighbourhood size distributions. In order to attempt an explanation of observed patterns, we develop a computational model of neighborhood dynamics where migration into and movement within the city are mediated by wealth. We find that the distribution of neighborhood sizes across 12 global cities follows exponential decay, and that this distribution is analytically consistent with empirically observed Zipf's Law for city sizes. We find that the model generates exponential decay in neighborhood size distributions for a range of parameter specifications. The use of a comparative wealth-based metric to assess the relative attractiveness of a neighborhood, combined with a stringent affordability threshold in mediating movement within the city are both found to be necessary conditions for the emergence of the exponential distribution.

Keywords: neighborhood, cities, power law, dynamics, urban, Zipf

1. Introduction

The distribution of city sizes in a country has been termed as an uncharacteristic regularity in economics (Gabaix, 1999; Krugman, 1996). Across many national contexts, city sizes are found to be distributed according to a power law, specifically the rank-size distribution of city sizes is said to obey Zipf's Law (Gabaix, 1999; Krugman, 1996; Luckstead & Devadoss, 2014; Mansury & Gulyas, 2007; Ioannides & Overman, 2003; Giesen & Südekum, 2011). If we were to rank cities based on their population size, Zipf’s Law posits that (Eqs. 1,2):

$$R_i = AP_i^{-\alpha} \qquad (1)$$

$$\ln R_i = \ln A - \alpha \ln P_i \qquad (2)$$

where R and P refer to the rank and population size, and A is a constant, then the estimated coefficient $\alpha \approx 1$. Large cross-country investigations also find some variance in the power law exponent across nations (Soo, 2005; Nitsch, 2005). A meta-study of 515 estimates from 29 studies of city size distribution from around the world finds the mean estimate of α to be 1.1 with two-thirds of the estimates ranging between 0.8 and 1.2 (Nitsch, 2005). Another cross-country analysis covering 75 countries was found to yield an average exponent of 1.1 (Soo, 2005). This regularity has sought to be explained using multiple theoretical models - in terms of the competing dynamics of new cities born at the rate of ν and existing cities growing at rate γ, yielding a power law with exponent $\alpha = \nu/\gamma$ (Steindl, 1968); using a stochastic growth model where migrants choose to form a new city with probability π and enter an existing city otherwise, resulting in $\alpha \approx 1$ when $\pi \approx 0$ (Simon, 1955); by assuming identical growth processes across city sizes (Gibrat's Law) resulting in Zipf's Law of city size distribution with $\alpha = 1$ (Gabaix, 1999); and by using agent based models where each agent (firm) makes decisions on its location based on the location's demographics, yielding Zipf's Law under certain conditions (Mansury & Gulyas, 2007).

In this work, we seek to fine-grain the scale of observation from the nation to the individual city, and study the distribution of neighborhood sizes across a city. To the best of our knowledge, this is a question that remains largely unexplored. Our work here specifically attempts to explore three questions. First, we seek to statistically characterize the distributions of neighborhood sizes across cities. Second, we build a computational model to isolate potential mechanisms underlying characteristic neighborhood size distributions. And third, we attempt to reconcile the findings on neighborhood size distributions with the Zipfian distribution of city sizes.

2. Empirical neighborhood size distributions

The notion of a neighborhood (much like the boundary of a city) defies strict definition. While there may be broad agreement on neighborhoods being geographically localised communities within a city, the exact boundaries of neighborhoods in any given city remain open to debate. Despite this lack of specific definition, it is important to recognize that deeply local processes involving both local communities and local administration have led to the emergence of neighborhood areas and their corresponding governance structures. While these context-specific conceptualizations of neighborhoods may not be consistent across nations, they do offer us a mechanism to explore the distribution of neighborhoods (that have emerged out of lived local experience) in cities across the world.

In this work, we study the distribution of neighborhood size across 12 global cities - Cape Town, Rio de Janeiro, Mumbai, New York, Moscow, Shanghai, London, Buenos Aires, Berlin, Dhaka, and Toronto. The choice of these cities was based on the availability of data at a sub-city (neighbourhood) level, and analysis of data from more cities would help further validate the veracity of these findings. Despite this limitation, the cities under analysis are spread across the world, ensuring diversity in historical contexts and socioeconomic outcomes. The notion of a neighborhood, as discussed earlier, is different across different cities - for instance, Mumbai's 97 wards each elect a Councillor, forming the level of government closest to the urban citizen, while New York City's Neighborhood Tabulation Areas (NTAs) were specifically created to be a summary level descriptor of the city's neighborhoods, offering a compromise between the broad strokes of the city's 59 districts and granularity of 2,168 census tracts, and Berlin's örststeiles are formally recognized localities for planning purposes, though not units of local government. Across the 12 cities under consideration (with average population 7.5 million), the average number of neighborhoods is 157, each with population 67,083. Appendix A provides the detailed data sources and neighborhood descriptions for all 12 cities. Table 1 presents a summary of neighborhood units used in this analysis.

City	Neighborhood type	Neighborhood count	Average neighborhood population
Cape Town	Suburb, Township	57	64,483
Rio de Janeiro	Bairro	159	40,103
Mumbai	Ward	97	128,272
New York City	Neighborhood Tabulation Area (NTA)	193	42,358
Moscow	Raiyon	123	86,042
Shanghai	Township-level Division	230	100,000
London	Ward	623	14,423
Buenos Aires	Barrio	48	57,826
Berlin	Ortsteile	96	35,034
Dhaka	Thana	41	161,711
Toronto	Neighborhood	172	13,800
Singapore	Planning Area	47	60,945

Table 1: *Neighborhood summary*

Unlike Zipf's Law, which appears to hold for distribution of city sizes, we do not find a consistent power law distribution of neighborhoods. Instead, we find that the distribution of neighborhoods across all cities under consideration is potentially best described by exponential decay, and this holds across the entire range of neighborhood sizes for each city (Fig. 1). What this essentially indicates is that, despite the large variations in neighborhood count and neighborhood sizes across cities (Table 1), the emergence of an exponential distribution of neighborhoods appears to be a consistent phenomenon in cities around the world.

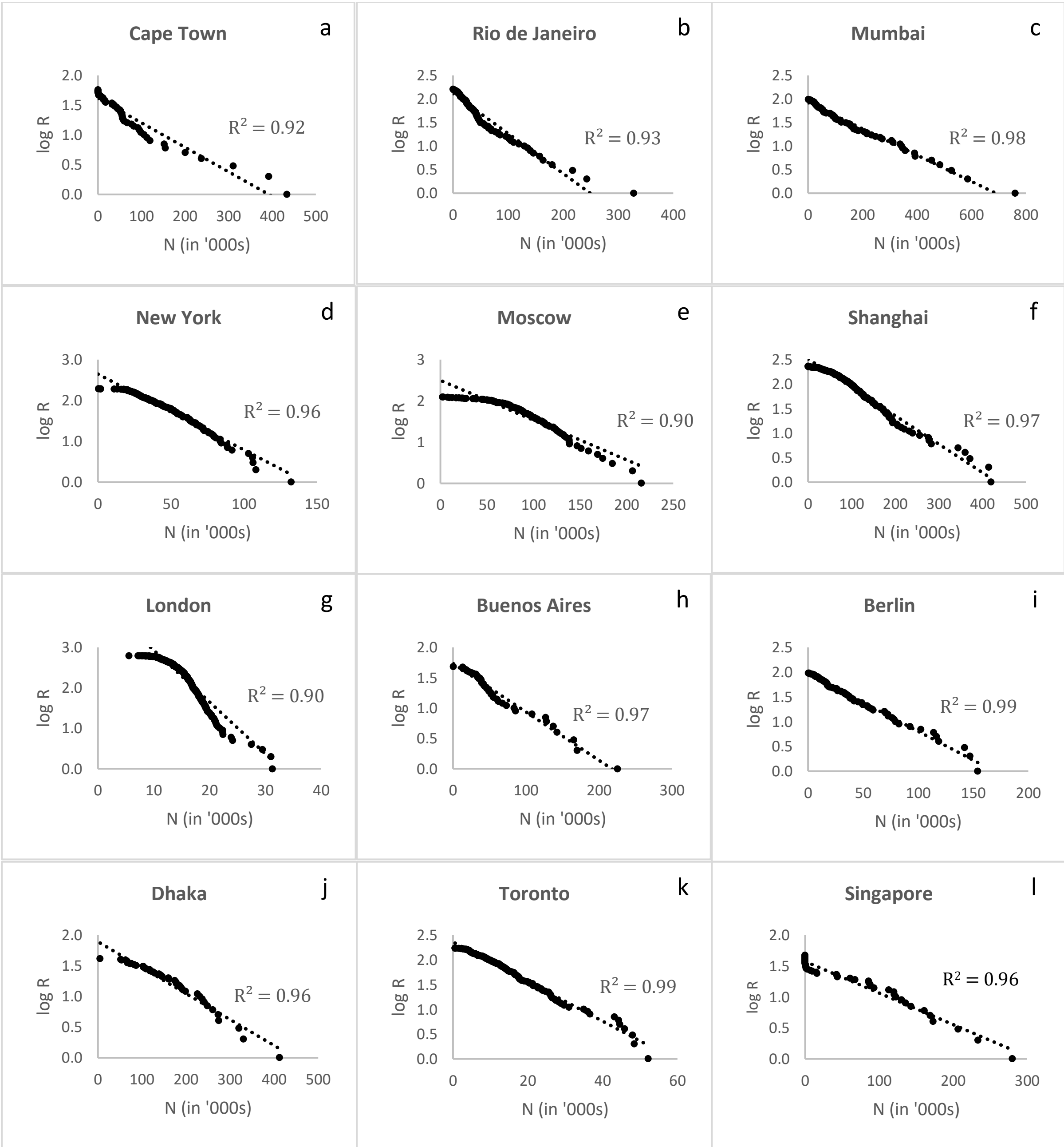

Figure 1. *log (Rank) v. Neighborhood Size*. a: Cape Town. b: Rio de Janeiro. c: Mumbai. d: New York. e: Moscow. f: Shanghai. g: London. h: Buenos Aires. i: Berlin. j: Dhaka. k: Toronto. l: Singapore. Across all cities, the distribution of neighborhood size shows exponential decay. Black Dots: Actual neighborhood size distributions. Dashed Black Line: Best fit line for exponential decay.

The available evidence suggests that the unequal distribution of neighborhood sizes in cities is reasonably characterised by exponential decay, which naturally leads us to ask why exponential decay describes neighborhood size distributions.

3. Model for neighborhood dynamics

In order to explore this question, we build a computational model of neighborhood evolution based on individual choices that are determined by the context of the neighborhoods that individuals inhabit, and attempt to isolate the general mechanisms that result in the emergence of exponential decay in neighborhood size distributions. The model we present here follows in the tradition of earlier models that explore the evolution of segregation and economic status of urban neighborhoods (Fossett, 2011; Benenson, Hatna, & Or, 2009; Zhang, 2004; Benard & Willer, 2007; Durrett & Yuan, 2014; Gargiulo, Gandica, & Carletti, 2017; Sahasranaman & Jensen, Ethnicity and wealth: The dynamics of dual segregation, 2018; Sahasranaman & Jensen, Cooperative dynamics of neighborhood economic status in cities, 2017), and belongs in the long tradition of threshold models going back to the original Schelling segregation model (Schelling, 1971). Given the efficacy of the Schelling family of models in simulating empirically observed patterns of racial, ethnic, status, and wealth-based segregations, this framework provides us a consistent basis to model neighbourhood demographic evolution and thereby to potentially generate one candidate explanation for the observed distribution patterns.

We consider a city of M neighborhoods with total population $P(0)$, where each neighborhood i ($i \in 1,2,\dots,M$) is initially composed of an equal number of agents, $P(0)/M$. Each agent is characterized by its wealth. The total wealth of neighborhood i is the sum of the wealths of all agents in i and denoted by $w_t(i)$.

This construction of a neighbourhood follows in the tradition of other metapopulation models where each neighbourhood is a location specified by a population at any given time and has no spatial geometry (Durrett & Yuan, 2014; Gargiulo, Gandica, & Carletti, 2017; Sahasranaman & Jensen, Cooperative dynamics of neighborhood economic status in cities, 2017); unlike the classic implementations of the Schelling model where each location in the lattice represents an individual agent, and neighbourhoods are constructed as explicit spatial geometries around a location such as von Neumann or Moore neighbourhoods in a two-dimensional lattice (Fossett, 2011; Zhang, 2004; Schelling, 1971).

While movement of agents within a city can be attributable to may reasons, it is unarguable that the ability to afford movement is a critical aspect of such a decision. If an agent is unable to afford to move into a neighbourhood, then she is unlikely to accomplish such a move however much she desires it. Previous empirical work, for instance, has shown that household income is the most important characteristic of the neighbourhood sorting process, much more so than aspects like education and employment status (Bolt & van Kempen, 2003; Clark & Ledwith, 2006; Hedman, Van Ham, & Manley, 2011). As in previous models we have built to explore neighbourhood dynamics, we make affordability the central basis for movement within the city (Sahasranaman & Jensen, Ethnicity and wealth: The dynamics of dual segregation, 2018; Sahasranaman & Jensen, Cooperative dynamics of neighborhood economic status in cities, 2017).

Each iteration of the model comprises migration into the city and movement within the city. First, agents attempt to migrate into the city, and the population attempting entry into the city is defined as a fraction r_{mig} of the city's extant population. However, the actual number of agents able to enter the city is determined by their individual wealths. If an incoming agent's wealth (w_m) is greater than the median wealth of a randomly chosen neighborhood j in the city (w_j^{med}), then the agent enters that neighborhood with probability 1. If not, the agent migration into the city is stochastic:

$$p_{ent} = \begin{cases} 1, & if\ w_m \geq w_j^{med} \\ e^{-\beta_m (w_j^{med} - w_m)}, & otherwise \end{cases}, \qquad (8)$$

where β_m is the calibrating factor for p_{ent}. Progressively decreasing β_m is reflective of increasing relaxation of the affordability condition. This could, for instance, be interpreted as public policy in the form of social housing or rental vouchers that enables households to move into neighborhoods that they are otherwise unable to afford. The base case value of β_m is chosen such that movement into neighborhoods in contravention of the wealth threshold condition is difficult (but not impossible), and is potentially reflective of real-world cities.

Once all agents have attempted migration into the city for a given iteration (time t), then the dynamics of movement within the city begin. At any given time t, P_t agents are randomly chosen to attempt movement within the city. The decision of a random agent (in neighborhood i) to move out of its location is based on the neighborhood's relative wealth. A random receiving neighborhood j is chosen, and the agent chooses to move with probability 1 if the median wealth of j (w_j^{med}) is greater than or equal to median wealth of i (w_i^{med}), and with probability 0 otherwise.

We argue that this a reasonable basis for determining an agent's choice to move, based on evidence of impacts that neighbourhoods have on socioeconomic outcomes of households (Chyn, 2018; Chetty & Hendren, The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects, 2018; Chetty & Hendren, The impacts of neighborhoods on intergenerational mobility II: County-level estimates, 2018). A study in Chicago found that moving children out of disadvantaged neighbourhoods resulted in lower school dropout rates, and also to greater likelihood of employment and higher wages as young adults, when compared to children who lived in disadvantaged neighbourhoods (Chyn, 2018). A study of 7 million households in the US showed that children's opportunities for economic mobility are impacted by their neighbourhoods – every extra year a child spent in a neighbourhood where resident incomes were higher, increased her own income (Chetty & Hendren, The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects, 2018). In general, neighbourhoods with lesser concentration of poverty and lower inequality, as well as better schools and lower crime rates produced better outcomes for children in poor families (Chetty & Hendren, The impacts of neighborhoods on intergenerational mobility II: County-level estimates, 2018). Neighbourhood income levels are also found to exert an effect on the health of individuals (Hou & Myles, 2005).

If the agent chooses to move, then the actual occurrence of the movement is mediated by its ability to afford the move. If agent wealth (w_a) is at least equal to the median wealth of j (w_j^{med}), then the agent moves with probability 1; the move becomes probabilistic otherwise:

$$p_{move} = \begin{cases} 1, & if\ w_a \geq w_j^{med} \\ e^{-\beta_m (w_j^{med} - w_a)}, & otherwise \end{cases} \quad (9)$$

β_m, the same parameter used to calibrate migration into the city, also calibrates movement within the city.

Table 2 lists the parameter values for the model simulations. The parameters are meant to reflect the range of urban dynamics resulting in neighborhood size distributions in real cities. We vary the rate of migration into the city as well as the correlation between wealth and status, with scenarios depicting both strong correlation between wealth and resident status and no correlation between wealth and resident status at all. Finally, we also study the sensitivity of outcomes to changes in the calibration parameter for migration and movement, β_m.

Parameter	Base case	Varying r_{mig}	Varying status-wealth correlation	Varying β_m
Number of neighborhoods, M	50	50	50	50
Initial population of agents, $P(0)$	2500	2500	2500	2500
Rate of population attempting entry per iteration, r_{mig}	0.005	0.01	0.005	0.005
Agent wealth distributions	$N(10,1)$ for residents;	$N(10,1)$ for residents;	$N(10,1)$ for residents;	$N(10,1)$ for residents;

	$N(7{,}1)$ for migrants	$N(7{,}1)$ for migrants	$N(10{,}1)$ for migrants	$N(7{,}1)$ for migrants
β_m	10	10	10	100; 5; 2; 1
Number of iterations	300	300	300	300

Table 2: *Model parameters*

4. Results and discussion:

We find that neighborhood distribution appears to be well approximated by an exponential distribution across a range of parameter specifications. When we consider the base case scenario, we find exponential decay in neighborhood size, which is in agreement with empirical observation (Fig. 1a). In the base case, $\beta_m = 10$, which indicates a non-zero probability of an agent (a) being able to move into neighborhood j in contravention of the wealth threshold of that neighborhood (w_j^{med}) (Eq. 9). We also find that doubling the fraction of population of migrants trying to enter the city at each iteration, $r_{mig} = 0.01$, yields an approximately exponential distribution as in the base case where $r_{mig} = 0.005$ (Fig. 1b). The nature of emergent dynamics therefore appears robust to changes in migration rates and system size. Finally, in the base case, migrant wealths are, on average, lower than resident wealths; so we test model outcomes by removing the correlation between resident status and wealth by drawing both resident and migrant wealths from the same distribution. We find that, even in this case, an exponential distribution results (Fig. 1c).

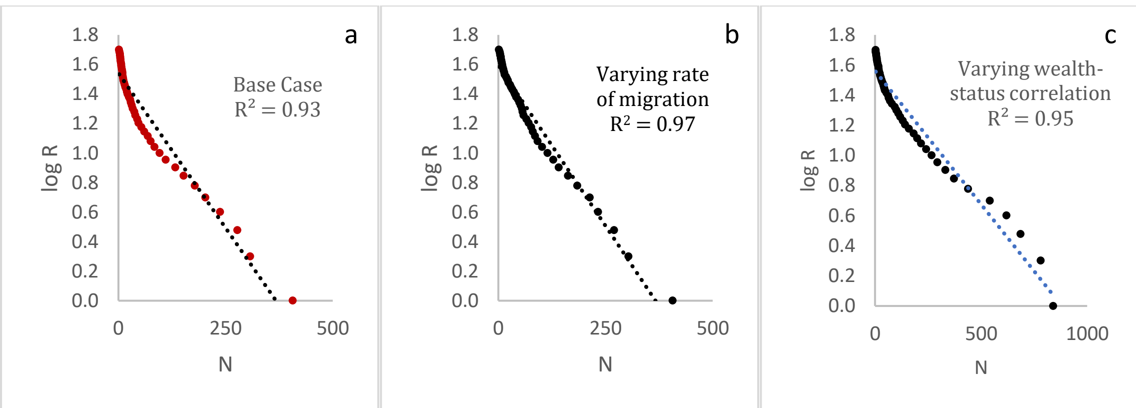

Figure 2. *Simulated plots of log (Rank) v. Neighborhood Size (N).* a: Base case. b: varying $r_{mig} = 0.01$, which is double the base case r_{mig} of 0.005. c: varying status-wealth correlation, ensuring no correlation between resident/migrant status and wealth (both drawn from the same wealth distribution), when compared to the base case when residents had higher wealth, on average. Across

all scenarios, we find that the emergence of exponential neighborhood size distributions is a robust result.

Overall, the emergence of an exponential distribution of neighborhood sizes is consistent across a range of parameter specifications, but all these scenarios assume wealth as the basis for migration into and movement within the city. While these results indicate that wealth, both as the criterion for agent choice of neighborhood and the mediator of agent movement based on affordability, is a sufficient condition for the emergence of exponential neighborhood distributions, the question of whether it is a necessary condition remains open, and once we explore next.

We test the efficacy of the wealth threshold criterion in generating exponential neighborhood size distributions: specifically, we explore how varying the stringency of the affordability condition, determined by β_m, impacts the emergence of neighborhood sizes. We vary β_m from 100 to 1 (taking values 100, 10, 5, 2, 1) (Table 2), and find that the emergence of the exponential distribution is consistent within a certain range of β_m, where only a low fraction amount of movement in contravention of neighborhood thresholds is possible (Figs. 3a, 3b). Beyond this range, neighborhood distributions are non-exponential (Figs. 3c, 3d, 3e). In order to quantify this range of β_m, we compute the threshold crossing ratio (TCR) as the ratio of number of times when an agent is able to successfully move into a neighborhood j despite its wealth (w_a) being lower than w_j^{med} (i.e. $w_a - w_j^{med} < 0$) to the total number of such attempts to move in contravention of the wealth threshold over the time of the dynamics (Fig. 3f). For $\beta_m \geq 10$, we find that $TCR \leq \sim 0.07$, and the resulting distribution is best approximated by an exponential (Figs. 3a, 3b). However, as we progressively relax the wealth threshold condition for $\beta_m \leq 5$ ($TCR > \sim 0.12$), we find that there is greater condensation of population into smaller fractions of neighborhoods, thus yielding closer approximations of potential power laws rather than exponentials (Figs. 3c, 3d, 3e, 3f). For instance, the top 10% of neighborhoods (by population) account for ~74% of the population when $\beta_m = 1$, when compared to only ~48% when $\beta_m = 10$. This is because, at lower β_m, larger fractions of poorer agents are able to move neighborhoods in contravention of the neighborhood wealth condition, resulting in population aggregation in a few neighborhoods.

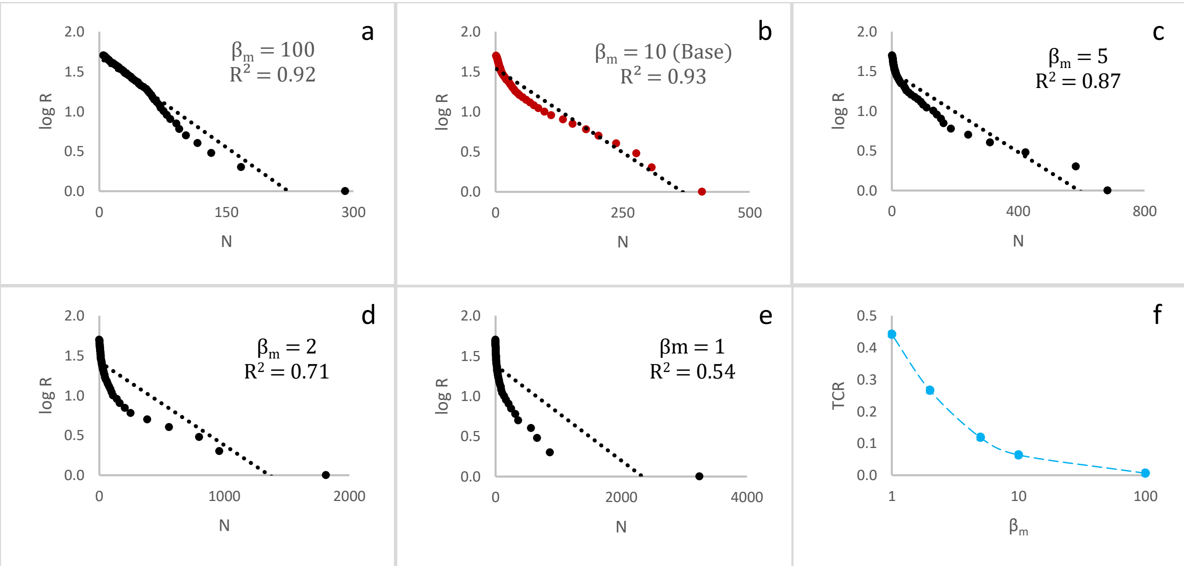

Figure 3. *Testing the necessity of wealth in generating exponential neighborhood distributions*. a: log R v. N for $\beta_m = 100$. b: log R v. N for $\beta_m = 10$. c: log R v. N for $\beta_m = 5$. d: log R v. N for $\beta_m = 2$. e: log R v. N for $\beta_m = 1$. As β_m decreases, the distribution of neighborhood sizes moves away from exponential decay indicating that a reasonably stringent affordability condition is operational in cities, resulting in exponential distribution of neighborhood populations. f: TCR v. β_m. As β_m decreases, TCR increases.

Overall, we find that while a small number of moves in contravention of wealth threshold conditions appears essential for an exponential distribution to emerge ($\beta_m = 100, TCR = \sim 0.01$), it is apparent that beyond a certain threshold ($\beta_m > 10, TCR > \sim 0.12$) the distribution of neighborhood size does not result in exponential decay. This finding confirms that the existence of a decision-making condition based on relative wealths of neighborhoods (w_i^{med} v. w_j^{med}), and an affordability condition mediating movement within the city ($\beta_m \geq 10, TCR < \sim 0.07$), are *both* necessary conditions in the model to yield an exponential distribution of neighborhood sizes. This finding is in concurrence with prior work on the importance of income in household decisions to move within cities (Bolt & van Kempen, 2003; Clark & Ledwith, 2006; Hedman, Van Ham, & Manley, 2011), as well as the influence of a neighbourhood's economic status on inter-generational social and economic outcomes of households (Chyn, 2018; Chetty & Hendren, The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects, 2018; Chetty & Hendren, The impacts of neighborhoods on intergenerational mobility II: County-level estimates, 2018).

The globally coupled nature of the dynamics under consideration makes exact analytical treatment very difficult, but we can undertake a simplified analytical exploration of the dynamics in a highly structured two neighborhood system. It is important to point out that this simplified illustration merely highlights the emergence of far from equal neighbourhood sizes under model dynamics, and not the emergence of exponential decay.

Consider a city system composed of two neighborhoods H_1 and H_2, populated by N agents. There are two kinds of agents – $\frac{N}{2}$ agents are of type A_1 with wealth w_1; and the remaining $\frac{N}{2}$ agents are of type A_2 with wealth w_2, such that $w_1 > w_2$. Initially, the agents are equally distributed across both neighborhoods, each with $\frac{N}{2}$ agents. At every point in time, each agent decides whether it wants to move from its current location i to j based on the wealth comparison between i and j: an agent moves only if $w_j^{med} \geq w_i^{med}$. Let us first consider a scenario where each neighborhood is populated by an equal number of A_1 and A_2 agents (Fig. 4a, initial). In this scenario, both neighborhoods have the same median wealth, $w_{H_1}^{med} = w_{H_2}^{med}$. Therefore, all agents are satisfied with their current locations and the initial configuration is an equilibrium configuration (Fig. 4a, final).

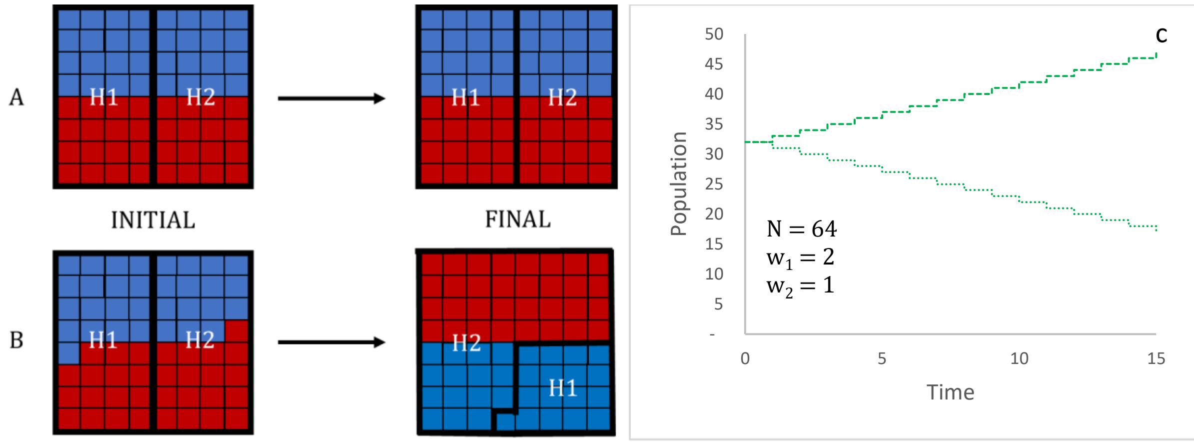

Figure 4. *Analytical description of dynamics in two-neighborhood system*. a: Initial state and Final equilibrium for a system ($N = 64, w_1 = 2, w_2 = 1$) where each neighborhood starts with an equal number of A_1 (red) and A_2 (blue) agents. Each square represents an individual agent and the thick black lines represent neighborhood boundaries. b: Initial and final configurations for a system which starts with an equal number of agents, but H_1 has one A_2 agent more than H_2 and H_2 has one A_1more agent than H_1. c: Evolution of total population over time in the two-neighborhood system. Evolution of total population in H_1 (dotted green) and population in H_2 (dashed green) shows that both neighborhoods have population of 32 to begin with, but final population of H_1 is 17, while that of H_2 is 47.

Now, let us perform the slightest perturbation of the initial condition and swap one A_1 agent from H_1 with one A_2 agent from H_2 (Fig. 4b, initial). The populations of the two cells remains equal at $\frac{N}{2}$, but the fraction of A_1 agents in H_1 is $\frac{\frac{N}{4}-1}{\frac{N}{2}} < 0.5$ and in H_2 is $\frac{\frac{N}{4}+1}{\frac{N}{2}} > 0.5$. Therefore, the median wealths of H_1 and H_2 are: $w_{H_1}^{med} = w_2$ and $w_{H_2}^{med} = w_1$ respectively. Given this configuration, all agents in H_2 are satisfied with their current location, but all agents in H_1 are unsatisfied in terms of neighborhood wealth comparison ($w_{H_2}^{med} > w_{H_1}^{med}$). Despite this dis-satisfaction with their current location, A_2 agents in H_1 are unable to move to H_2 because their wealths (w_2) are lower than the median wealth of H_2 ($w_{H_2}^{med} = w_1 > w_2$). Therefore, the final equilibrium in this two-neighborhood system involves all A_1 agents in H_1 moving to H_2, because their wealths w_1 are equal to the median wealth of H_2, which remains at $w_{H_2}^{med} = w_1$ for the duration of the dynamics. This results in an unequal distribution of population across the two neighborhoods, with H_1's final population at $\frac{N}{4} + 1$ and H_2's at

$\frac{3N}{4} - 1$. Figure 4c illustrates the evolution of a two-neighborhood system for specific values of N, w_1, w_2 and its final equilibrium in terms of total population. What this simple analytical illustration demonstrates is that even the mildest perturbation in the initial condition in a highly structured two-neighborhood system with equal populations at the outset, results in an equilibrium with far from equal population sizes across neighborhoods.

5. From exponential decay of neighborhoods to Zipf's Law in cities

Now that we have proposed a candidate mechanism to explain the emergence of exponential decay in neighbourhood sizes, we turn to explore how this distribution of neighbourhoods relates to the Zipfian distribution of city sizes.

We begin from our finding that each city has, approximately, an exponential distribution of neighborhood sizes (c). The probability of a neighborhood with size $c = C$ is:

$$P\{c = C\} = \frac{1}{\lambda} e^{-C/\lambda} \quad (1)$$

Let N denote the number of neighborhoods in a city. The size of a city (s) is the sum of its neighborhood sizes:

$$s = \sum_{i=1}^{N} c_i \quad (2)$$

Given that the sum of N exponentially distributed variables is a gamma distribution, the probability of a city of size $s = S$ is:

$$P\{s = S\} = \frac{\lambda^N S^{N-1} e^{-\lambda S}}{\Gamma(N)} \quad (3)$$

Now, consider a set of cities characterised by a specific set of values of λ and N. Assume that the probability that a city picked at random has parameters λ and N is given by $P(N, \lambda)$. Therefore, the probability of a randomly selected city being of size S is:

$$P\{s = S\} = \sum_N \sum_\lambda P(N, \lambda) \, \frac{\lambda^N S^{N-1} e^{-\lambda S}}{\Gamma(N)} \quad (4)$$

Assuming that N and λ are real positive numbers, we can replace the summations by integrals over the real axis. Therefore:

$$P\{s = S\} = \int_0^\infty dN \int_0^\infty d\lambda \ P(N, \lambda) \ \frac{\lambda^N S^{N-1} e^{-\lambda S}}{\Gamma(N)} \qquad (5)$$

Replacing λ with x/S, we get:

$$P\{s = S\} = \frac{1}{S^2} \int_0^\infty dN \int_0^\infty dx \ P(N, x/S) \ \frac{x^N e^{-x}}{\Gamma(N)} \qquad (6)$$

The dependence on S is only through the term $P(N, x/S)$, which if it is a weak dependence, yields the approximation:

$$P\{s = S\} \sim S^{-2} \qquad (7)$$

This is equivalent to Zipf's Law for city size distributions. Indeed, the extent of dependence on the $P(N, x/S)$ term possibly explains the variation of the power law exponent for city size distributions around the exact value of the Zipf exponent ($\alpha = 1$), as observed in large meta-studies for city systems around the world (Soo, 2005; Nitsch, 2005). Eqs. 1-7 therefore analytically demonstrate that the exponential decay observed in neighborhood sizes is consistent with Zipfian distribution of city sizes.

6. Conclusion:

We study neighborhoods across a set of 12 global cities and find that the distribution of neighborhood sizes follows exponential decay across all cities under consideration. In order to explore the emergence of the exponential distribution of neighborhood size, we build a computational model of wealth-based neighborhood dynamics. In this model, agents assess their satisfaction with their extant neighborhoods by using a simple wealth-based metric which compares their neighborhood's median wealth with that of a randomly chosen neighborhood in the city. If satisfied, agents stay back in their current neighborhood and if not, they attempt to move. Movement to another neighborhood is mediated by an affordability condition. Using this simple set up, we find that the dynamics yield exponential neighborhood size distributions, in concordance with empirical observations. Most importantly, we find that using the wealth condition as the basis for decision making and movement within the city, in conjunction with the affordability condition, is essential for the emergence of exponential decay.

While a closed form analytical description eludes us due to the complexity of the dynamics, we construct a simple, highly structured two-neighborhood system to illustrate the emergence of unequal neighborhood sizes.

Finally, we also explore the question of whether exponential neighborhood sizes are consistent with Zipfian city size distributions. We demonstrate analytically that city populations, composed of exponentially decaying neighborhood sizes, can be distributed as a power law with an exponent in the region of 1.

Abbreviations:
NTA: Neighborhood Tabulation Area
TCR: Threshold Crossing Ratio

Declarations:

Availability of data and materials: All data generated or analysed during this study are included in this published article and its supplementary information files (Appendix A).

Competing Interests: The authors declare that they have no competing interests.

Funding: The authors received no funding for this work.

Authors' contributions: AS conceptualized the research, programmed the simulations, and wrote the draft manuscript. AS and HJJ designed the model, wrote the mathematical description, reviewed and edited the manuscript.

Acknowledgements: AS acknowledges the feedback from participants at the Urban Complex Systems workshop at the Conference for Complex Systems 2020.

References

Benard, S., & Willer, R. (2007). A Wealth and Status-Based Model of Residential Segregation. *J. Math. Sociol., 31*, 149–174.

Benenson, I., Hatna, E., & Or, E. (2009). From Schelling to Spatially Explicit Modeling of Urban Ethnic and Economic Residential Dynamics. *Sociol. Meth. and Res. , 37*(4), 463–497.

Bolt, G., & van Kempen, R. (2003). Escaping Poverty Neighbourhoods in the Netherlands. *Housing, Theory and Society, 20*, 209-222.

Chetty, R., & Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects. *The Quarterly Journal of Economics, 133*(3), 1107-1162.

Chetty, R., & Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility II: County-level estimates. *The Quarterly Journal of Economics, 133*(3), 1163-1228.

Chyn, E. (2018). Moved to opportunity: The long-run effects of public housing demolition on children. *American Economic Review, 108*(10), 3028-3056.

Clark, W., & Ledwith, V. (2006). Mobility, housing stress and neighbourhood contexts: evidence from Los Angeles. *Environment and Planning A, 38*, 1077-1093.

Durrett, R., & Yuan, Z. (2014). Exact solution for a metapopulation version of Schelling's model. *PNAS, 111*(39), 14036–14041.

Fossett, M. (2011). Generative Models of Segregation: Investigating Model-Generated Patterns of Residential Segregation by Ethnicity and Socioeconomic Status. *J Math Sociol, 35*(1-3), 114–145.

Gabaix, X. (1999). Zipf's law for cities: An explanation. *The Quarterly Journal of Economics*, 114(3): 739-767.

Gargiulo, F., Gandica, Y., & Carletti, T. (2017). Emergent dense suburbs in a Schelling metapopulation model. *Adv. in Complex Sys. , 20*(1), 1750001.

Giesen, K., & Südekum, J. (2011). Zipf's law for cities in the regions and the country. *Journal of Economic Geography, 11*(4), 667-686.

Hedman, L., Van Ham, M., & Manley, D. (2011). Neighbourhood choice and neighbourhood reproduction. *Environment and Planning A, 43*(6), 1381-1399.

Hou, F., & Myles, J. (2005). Neighbourhood inequality, neighbourhood affluence and population health. *Social science & medicine, 60*(7), 1557-1569.

Ioannides, Y., & Overman, H. (2003). Zipf's law for cities: an empirical examination. *Regional science and urban economics, 33*(2), 127-137.

Krugman, P. (1996). Confronting the mystery of urban hierarchy. *Journal of the Japanese and International economies*, 10(4): 399-418 .

Luckstead, J., & Devadoss, S. (2014). A comparison of city size distributions for China and India from 1950 to 2010. *Economics Letters, 124*, 290-295.

Mansury, Y., & Gulyas, L. (2007). The emergence of Zipf's Law in a system of cities: An agent-based simulation approach. *Journal of Economic Dynamics & Control, 31*, 2438–2460.

Nitsch, V. (2005). Zipf zipped. *Journal of Urban Economics, 57*(1), 86-100.

Rosen, K., & Resnick, M. (1980). The size distribution of cities: An examination of the Pareto law and primacy. *Journal of Urban Economics, 8*(2), 165-186.

Sahasranaman, A., & Jensen, H. (2017). Cooperative dynamics of neighborhood economic status in cities. *Plos One, 12*(8), e0183468.

Sahasranaman, A., & Jensen, H. (2018). Ethnicity and wealth: The dynamics of dual segregation. *Plos One, 13*(10), e0204307.

Schelling, T. (1971). Dynamic models of segregation. *Journal of mathematical sociology, 1*(2), 143-186.

Simon, H. (1955). On a Class of Skew Distribution Functions. *Biometrika, XLII*, 425-440.

Soo, K. (2005). Zipf's Law for cities: a cross country investigation. *Regional Science and Urban Economics, 3*, 239-263.

Steindl, J. (1968). Size Distributions in Economics. In *International Encyclopedia of the Social Sciences, Vol. 14.* New York: Macmillan and the Free Press.

Zhang, J. (2004). A dynamic model of residential segregation. *J. Math. Sociol., 28*, 147–170.

APPENDIX A:

This appendix describes the sources of data for the analysis on neighborhood size distribution.

1. Cape Town: Population data for Cape Town's main places is available from Census of South Africa 2011 data at https://census2011.adrianfrith.com/place/199. Detailed data on places is also available from Statistics South Africa at http://www.statssa.gov.za/?page_id=993&id=city-of-cape-town-municipality. These main places encompass suburbs, towns, and townships that comprise the city of Cape Town.

2. Rio de Janeiro: Population data at the level of Rio's bairros is available from the Brazilian census IBGE of 2010 at https://cidades.ibge.gov.br/brasil/rj/rio-de-janeiro/pesquisa/19/29761, which is the basis for Wikipedia's pages on Rio's bairros at https://pt.wikipedia.org/wiki/Lista_de_bairros_da_cidade_do_Rio_de_Janeiro. The municipal administration of Rio is subdivided into sub-prefectures, which are further sub-divided into 33 administrative regions, and finally into bairros.

3. Mumbai: Ward level population data for Mumbai was obtained from the Census of India 2011 at https://censusindia.gov.in/pca/pcadata/Houselisting-housing-Maha.html (choice: Mumbai). The ward is the lowest level of urban local government, with residents of a ward electing a Councillor, who is the elected official closest to the citizen and responsible for local issues.

4. New York City: Neighborhood Tabulation Area (NTA) level population data for NYC is available from US Census data for 2010 at https://data.cityofnewyork.us/City-Government/Census-Demographics-at-the-Neighborhood-Tabulation/rnsn-acs2/data. The NTAs were specifically created to be a summary level descriptor of the city's neighborhoods, offering a compromise between the broad strokes of the city's 59 districts and granularity of 2,168 census tracts.

5. Moscow: Data for the raiyons (districts) of Moscow is available from the Russian census of 2002 at http://www.perepis2002.ru/index.html?id=87 and presented in Wikipedia at https://en.wikipedia.org/wiki/Administrative_divisions_of_Moscow. The raiyons are municipal or local government entities under its 12 administrative okrugs within Moscow.

6. Shanghai: Data for township-level divisions of Shanghai are available from China's population census 2010, presented in Wikipedia at https://en.wikipedia.org/wiki/List_of_township-level_divisions_of_Shanghai#cite_note-5. Shanghai had 19 districts, under which the township-level divisions comprise 104 districts, 107 towns, two townships, and some special township-level divisions.

7. London: Ward level population data for London is provided by the Greater London Authority and available at https://data.london.gov.uk/dataset/land-area-and-population-density-ward-and-borough for 2011 through 2050 (projections). We use 2018 data for the analysis. The wards are the lowest levels of local government and are contained within the 33 boroughs of Greater London.

8. Buenos Aires: Population data for Buenos Aires' formal barrios (neighborhoods) is available from Argentina's census of 2010, tabulated in Wikipedia at https://en.wikipedia.org/wiki/Neighbourhoods_of_Buenos_Aires. These barrios are grouped under 15 communes, which are units of decentralized local government in Buenos Aires.

9. Berlin: Population data for the ortsteiles (localities) as of 2008 is available from Statistik Berlin-Brandenburg at https://www.statistik-berlin-brandenburg.de/publikationen/Stat_Berichte/2008/SB_A1-5_h2-07_BEneu.pdf, and compiled in Wikipedia at https://en.wikipedia.org/wiki/Boroughs_and_neighborhoods_of_Berlin. Berlin is constituted of 12 boroughs, which are in turn composed of officially recognized ortsteiles. The ortsteiles are not units of local government, but recognized for planning purposes.

10. Dhaka: Population of Dhaka's thanas are obtained from the Bangladesh Population and Housing Census 2011 available at https://web.archive.org/web/20151208044832/http://www.bbs.gov.bd/WebTestApplication/userfiles/Image/National%20Reports/Union%20Statistics.pdf, and also Banglapedia - the national encyclopaedia of Bangladesh by scholars at http://en.banglapedia.org/index.php?title=Main_Page. The thana started as a unit of police administration and was upgraded into a unit of municipal administration.

11. Toronto: Population data on Toronto's neighborhoods is constructed at https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods from Canada's census of 2006 https://www12.statcan.gc.ca/census-recensement/2006/rt-td/index-eng.cfm. Neighborhood boundaries are approximated to the nearest census tract from the census data.

12. Singapore: Population for Singapore's planning areas is obtained from Singapore's Open Government Data portal at https://data.gov.sg/dataset/singapore-residents-by-planning-area-subzone-age-group-and-sex?resource_id=ad854cc4-f9a3-4208-a9e5-cb8d7fb0a76c. Planning areas are the main urban planning divisions of the city.