Zipf's law from scale-free geometry

Henry W. Lin[1] and Abraham Loeb[2]

[1]*Harvard College, Cambridge, Massachusetts 02138, USA*

[2]*Institute for Theory & Computation, Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA*

(Received 19 April 2015; revised manuscript received 9 September 2015; published 7 March 2016)

The spatial distribution of people exhibits clustering across a wide range of scales, from household ($\sim10^{-2}$ km) to continental ($\sim10^{4}$ km) scales. Empirical data indicate simple power-law scalings for the size distribution of cities (known as Zipf's law) and the population density fluctuations as a function of scale. Using techniques from random field theory and statistical physics, we show that these power laws are fundamentally a consequence of the scale-free spatial clustering of human populations and the fact that humans inhabit a two-dimensional surface. In this sense, the symmetries of scale invariance in two spatial dimensions are intimately connected to urban sociology. We test our theory by empirically measuring the power spectrum of population density fluctuations and show that the logarithmic slope $\alpha = 2.04 \pm 0.09$, in excellent agreement with our theoretical prediction $\alpha = 2$. The model enables the analytic computation of many new predictions by importing the mathematical formalism of random fields.

DOI: 10.1103/PhysRevE.93.032306

I. INTRODUCTION

Human populations exhibit remarkably simple properties given the complexity of socioeconomic interactions between humans and their environments [1]. One such example is the well known *Zipf's law* [2] for cities: the rank of a city is inversely proportional to the number of people who live in the city. If the most populous city in the United States has a population of $N_{\max,\mathrm{US}} \sim 8 \times 10^6$, the second most populous city will have a population of $\frac{1}{2}N_{\max,\mathrm{US}} \sim 4 \times 10^6$, the third $\frac{1}{3}N_{\max,\mathrm{US}} \sim 2.7 \times 10^6$, and so forth. This simple relation fits empirical data extremely well [3,4]. A mathematically equivalent formulation of Zipf's law is that the underlying distribution of cities follows a power law [5]; namely, the probability that a city has a population N scales as $1/N^2$.

The remarkable simplicity and empirical success of Zipf's law have attracted significant theoretical attention and debate [3,6,7], though there is no consensus on the origin of Zipf's law. Existing work treats cities as the fundamental entities of the theory, with population as a property of each city. For example, Gibrat's law applied to cities [3,8,9], which states that the fractional growth rate of a city is independent of its population, will drive the distribution of city populations to a log-normal distribution. The tail of the log-normal distribution then gives rise to Zipf's law.

Our approach is conceptually different: we treat the population density as the fundamental quantity, thinking of cities as objects that form when the population density exceeds a critical threshold. The situation is therefore conceptually and mathematically analogous to the formation of galaxies in the universe, where nonlinear gravitational collapse occurs when the matter density exceeds some critical value. Our conceptual advance here is also a practical one, since we can apply the mathematical tools developed for analyzing random fields [10] to the problem at hand.

Before proceeding with a technical derivation of our results, let us briefly summarize them. The starting point is to model human population density as a random function of spatial position. A function of spatial position is a *field*, and thus human population density will be modeled as a *random field* (for a review of relevant topics in random fields, see [11,12] and especially [13]). To lowest order, a single random variable in elementary statistics is characterized by a mean and a variance. A random field may be regarded as a higher-dimensional generalization of a single random variable. By analogy, a random field is characterized by a mean and a *power spectrum*, which can be thought of as a generalization of variance. The power spectrum gives the amount of fluctuations of the field as a function of scale. To derive the form of the power spectrum for human population density, we invoke scale-invariant random growth, similar in spirit to Gibrat's law.

We move on to derive Zipf's law. Our derivations involve the simple assumption that some cities emerge above some critical population density threshold. To count the number of cities in our model, one must answer the following mathematical question: given a random field characterized by a power spectrum, how often does the random field take on values greater than a certain threshold? This is a frequently asked question in the context of cosmology, and the Press-Schechter (PS) formalism allows us to analytically compute the answer.

We demonstrate that our derivation of Zipf's law is more general than the motivating random growth model; we argue that the only key ingredient is scale invariance in two spatial dimensions. In other words, whereas previous work tends to focus on how Zipf's law emerges from concrete models, we argue that Zipf's law naturally occurs in a very large class of statistical models. In the language of statistical physics [12], the existence of Zipf's law is a function only of the universality class of the statistical model; it is independent of the "microscopic" details of the system's dynamics, which are undoubtedly complex in the case of human populations.

II. DERIVATION OF ZIPF'S LAW

We now proceed with the detailed derivation. To start, consider the human population density ρ as a function on $\mathbb{R}^2$, the two-dimensional (2D) Euclidean plane. Since we will be interested in regions much smaller in size than the radius of the Earth, we will ignore the effects of curvature. The fluctuations relative to the average population density $\delta(\mathbf{x}) \equiv \{[\rho(\mathbf{x})/\bar{\rho}] - 1\}$ can be expanded in Fourier modes,

$$\delta(\mathbf{x}) = \frac{1}{2\pi} \int d^2k \, \delta_{\mathbf{k}} \, e^{-i\mathbf{k}\cdot\mathbf{x}}.$$

Up to a conventional normalization factor of 2π, this equation simply rewrites the population fluctuations as a sum of plane waves $e^{-i\mathbf{k}\cdot\mathbf{x}}$, each weighted by a factor $\delta_{\mathbf{k}}$. Since the left-hand side is a random variable, the right-hand side must also be a random variable; since every term except for $\delta_{\mathbf{k}}$ on the right-hand side is manifestly deterministic, $\delta_{\mathbf{k}}$ must be a continuum of random variables, with one random variable for each wave vector $\mathbf{k}$. Just as an ordinary random variable is characterized by a variance, each $\delta_{\mathbf{k}}$ is characterized by a number $P(\mathbf{k})$ called the power spectrum, which is defined as

$$\langle \delta_{\mathbf{k}} \, \delta^*_{\mathbf{k}'} \rangle = (2\pi)^2 \delta_D^2(\mathbf{k} - \mathbf{k}') P(k), \tag{1}$$

where δ_D is the Dirac delta function (not to be confused with the fractional overdensity $\delta(\mathbf{x})$. With the assumption of rotational symmetry, the power spectrum becomes a function only of magnitude $P(\mathbf{k}) = P(k)$. Equation (1) makes precise the statement that the power spectrum $P(k)$ quantifies the amount of statistical fluctuations associated with a given frequency k.

It is conventional to define a dimensionless power spectrum in the number density $\Delta^2(k) \equiv k^2 P(k)/(2\pi)$, which represents the typical (squared) fractional overdensity of people $(\delta\rho/\rho)^2$ on the spatial scale $\sim 1/k$. To make further progress, we must fix the functional form of $\Delta(k)$ by some theoretical principle. To this end, consider an overdensity of size $\sim 1/k$. At a discrete time step, this overdensity might grow or shrink in spatial coverage. As a concrete example, consider a collection of farms (with a characteristic population density of a few people per typical farm area) in otherwise relatively uninhabited countryside. At each time step, a farm could be added or destroyed. In this way, our unifying principle of random walkers is conceptually similar to previous work on the random growth of firms [14]. Therefore, the spatial size of the overdensity might grow or shrink, while $\delta\rho/\rho$ (a number associated with farms) will be held constant. More precisely, we define a monotonically decreasing function $X(k)$ such that $\lim_{k\to\infty} X = 0$, which quantifies the spatial extent of an overdensity. This function might represent the area of the overdensity $X(k) \propto 1/k^2$ or its perimeter $X(k) \propto 1/k$, but our derivation will not depend on the detailed form of X. We can then perform a change of variables and view $\Delta(k)$ as a function of X: $\Delta(X(k)) = \Delta(k)$. The unifying principle is that all overdensities can grow or shrink spatially, executing a random walk in X. This process can continue until the overdensity disappears ($X = 0$), or the overdensity takes up some maximum $X_{\max}$, where $X_{\max} \equiv X(k_{\min})$ is set by the continental length scale $\sim 1/k_{\min}$. For a large ensemble of overdensities, this is a diffusionlike process with reflecting boundary conditions obeying

$$\frac{\partial \Delta}{\partial t} = D \frac{\partial^2 \Delta}{\partial X^2} \tag{2}$$

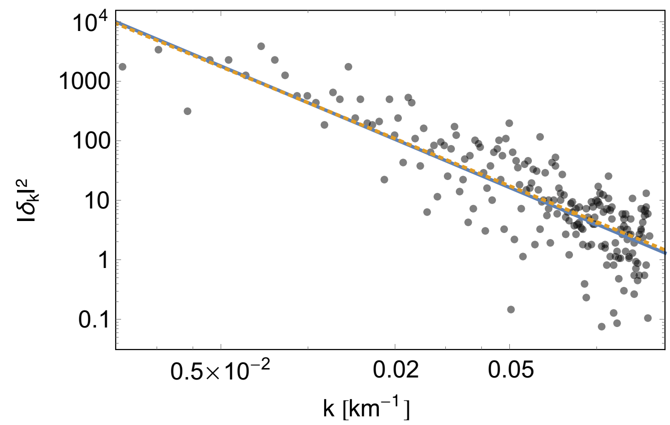

FIG. 1. Empirically measured power spectrum $P(k) \propto |\delta_k|^2 \sim k^{-\alpha}$ of population density fluctuations as a function of the spatial wave number k. The best-fit slope $\alpha = -2.04 \pm 0.09$ (solid blue line) is virtually indistinguishable from the predicted slope $\alpha = -2$ (dashed orange line). The data were obtained by taking the diagonal entries (to avoid anisotropy from rectangular gridding) of a discrete Fourier transform of a 1000×1000 arcmin^2 map of the population density of a section of the continental United States. The area was selected to minimize artifacts due to boundary conditions defined by lakes and oceans.

with some diffusion constant D. We are interested in only the late-time behavior of Eq. (2). Any initial conditions will relax to the steady-state solution $\Delta(X) \to$ constant for $0 \leqslant X \leqslant X_{\max}$ on a time scale $T_{\text{relax}} \sim X_{\max}^2/D$. We intuitively expect T_{relax} to be reasonably short, since the geographic mobility time scale of $\sim$5 y (in the United States, $\sim$35% of people change residences within 5 y [15]) is considerably shorter than, say, the population growth time scale $\sim$30 y set by the typical age of parenthood. Any initial conditions set by antiquity or perturbations to the system (e.g., catastrophic events that displace many people) should be quickly erased. We therefore predict that on sufficiently long time scales

$$P(k) \propto k^{-2}. \tag{3}$$

We test this prediction in Fig. 1 against publicly available data from the Center for International Earth Science Information Network (CIESIN) and Centro Internacional de Agricultura Tropical (CIAT) [16]. We find the best-fit slope $P(k) \propto k^{-\alpha}$ to be $\alpha = 2.04 \pm 0.09$, where we have reported the $\pm 1\sigma$ uncertainties. Our theoretical prediction is therefore in excellent agreement with observations across a broad range of spatial scales, from a few km to $\sim 10^3$ km.

Before further developing the theory, a more intuitive derivation of $P(k) \sim k^{-2}$ is worth mentioning. Over a large range of length scales our model is scale-free, implying $P(k) \sim k^{-\alpha}$ for some α. In d spatial dimensions, the left-hand side of Eq. (1) has units of k^{-2d}, the Dirac δ function has units of k^{-d}, so $P(k)$ should have units of k^{-d}. Since there are no other dimensional parameters relevant to our theory (the diffusion constant has units of $[X]^2 T^{-1}$, but there are no other constants with units of time T), we must have $\alpha = d = 2$ in two spatial dimensions. In this sense, geometry and scale invariance uniquely determine the slope of the power spectrum.

In fact, this simple argument demonstrates that $P(k) \propto k^{-2}$ is a universal feature of 2D models that have no parameters with units of length to some power. Effective field theory, a powerful technique for studying any statistical physics system, can be used to further sharpen this statement; this is done in Appendix D. The derivation in Appendix D provides perhaps the most rigorous way of justifying the statement that $P(k) \propto k^{-2}$ is a generic property of models with scale invariance in two dimensions, since effective field theory should capture the equilibrium properties of any statistical model on scales smaller than the system scale but sufficiently large such that densities can be approximated by smooth functions.

With a power spectrum $P(k)$ in hand, it is possible to calculate the number of cities as a function of their population N. We picture cities of area A as discrete objects which form when the population density as a function of spatial coordinates $\rho(\mathbf{x})$, or equivalently $\delta(\mathbf{x})$, averaged over an area A surpasses a critical threshold δ_C. In other words, we choose the surface area A such that the total integrated population $N = \int_{x\in A} \rho(\mathbf{x})\, d^2x = \rho_C A$, where the critical density $\rho_C = \bar{\rho}(1+\delta_C)$.

This is shown pictorially in Fig. 2. (This assumption can be relaxed, allowing for the average population density of a city to vary systematically with size. In our model, this corresponds to a critical threshold that varies with A. In this case, the excursion set formalism can be used with a moving barrier [17]. We will ignore this subtlety, since Zipf's law is still obtained in the limit that $\delta_C \ll \sigma$. Also, since the numerical value of the threshold is not fixed, we could also consider the case where the threshold varies by country. Again, Zipf's law would be obtained for each country.)

The counting of cities is now a well-posed question. Computationally, one could find the number distribution of cities with the following algorithm. Generate via a Monte Carlo procedure many realizations of the random field with mean 0 and power spectrum $P(k) = P_0 k^{-2}$. Find the regions where the random field exceeds a certain threshold. Measure the size of each region, and multiply the area of each region by the population density threshold; define this to be the population of each city. Repeat for many Monte Carlo iterations, and then make a histogram of the size distributions of each region. One can verify numerically that the resulting number distribution $n(N)$ will scale approximately as

$$n(N) \propto N^{-2}, \tag{4}$$

where N is the population of the city and $n(N)$ is the number density of cities of size N. However, we can in fact show *analytically* that the number distribution takes this form using the Press-Schechter formalism [10], traditionally used in the context of cosmology to predict the abundance of gravitationally bound objects given a power spectrum of the fluctuations in the cosmic matter density. However, we emphasize that the formalism is in essence a purely statistical one, which does not require or employ any facts from cosmology. The excursion set formalism [18] provides a more rigorous derivation, but the PS formalism has the benefit of simplicity. The end result is identical in either case. We provide a self-contained proof of Eq. (4) in Appendix A.

By integrating Eq. (4) with respect to N, we find that the number of cities above a certain population threshold scales inversely with the population threshold. This statement is equivalent to Zipf's law: the rank of a city is inversely proportional to its size.

III. CONCLUSION

In summary, we have presented a derivation of Zipf's law and successfully predicted the power spectrum $P(k)$ of population density fluctuations in the continental United States. These derivations stemmed from two fundamental ingredients: scal invariance and 2D geometry. Remarkably, there is a wide range of possible models and an even wider range of initial conditions to which our results are insensitive. One such model involves random walks of the sizes of clusters of people on all scales, which can be viewed as a vast generalization of Gibrat's law. However, as we have emphasized, even this generalization is still a relatively specific example in the class of all models which will lead to Zipf's law. This shows that the origin of these laws is fundamentally the scale-free nature of clustering in human populations. This is an appealing feature, enabling us to forgo any fine-tuning arguments in explaining the empirical data.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous referees for useful discussions. This work was supported in part by NSF Grant No. AST-1312034.

APPENDIX A: THE PRESS-SCHECHTER FORMALISM

The Press-Schechter formalism (for a pedagogical overview of the PS formalism and its generalizations, see Sec. 3.4 of [19]) allows us to answer the well-posed question: given a random field with an associated power spectrum $P(k)$, how often does it exceed the threshold? More specifically, suppose there is a class of objects (e.g., cities) that form when the population density exceeds a certain threshold $\rho(x) > \rho_{\text{threshold}}$. Furthermore, let the size R of each object be defined as the maximum radius R such that the average population density $\rho_{\text{circle,R}}$ within a circle of radius R centered on the object is given by

$$\rho_{\text{circle,R}} = \rho_{\text{threshold}}. \tag{A1}$$

It is conventional to define a smoothed density field

$$\delta_A(\mathbf{x}) = \int d^2k\, W_A(k)\, \delta_{\mathbf{k}}\, e^{-i\mathbf{k}\cdot\mathbf{x}}/(2\pi)^2, \tag{A2}$$

where the low-pass window function $W_A(k) = 1$ if $k \leqslant 1/\sqrt{A}$ and $W_A(k) = 0$ otherwise. This smoothed field is simply the original field $\delta(x)$ with the high-frequency fluctuations subtracted out, leaving behind the slowly varying components.

For a fixed $\mathbf{x}$, $\delta_A(\mathbf{x})$ is a random variable with probability distribution $p_A(\delta_A)$. The key insight of the PS formalism is to identify the fraction f_A of people living in cities of area A or larger with the cumulative probability $f = 2\int_{\delta_C}^{\infty} p_A\, d\delta$. This is illustrated in Fig. 2. To make further progress, we must assume something about the functional form of p_A. The conventional

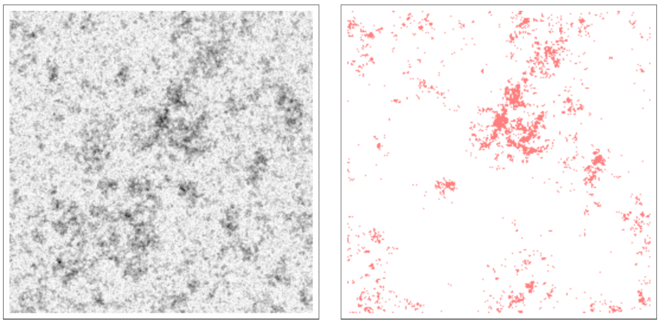

FIG. 2. Schematic illustration of our approach. On the left, a simulated population density map with a power spectrum $P(k) \propto k^{-2}$ is displayed. Darker pixels indicate higher population densities. On the right, we select and color pink all pixels above a certain population density threshold from the simulated map on the left. In our formalism, a city is identified with each pink cluster, appropriately smoothed on the length scale of the cluster. Scale invariance implies that cities of all sizes appear on the map, as confirmed by visual inspection. In our formalism, the statistical size distribution of pink clusters gives us the population distribution of cities, which agrees with Zipf's law.

PS formalism assumes that p_A is a Gaussian with mean 0 and variance $\sigma^2(A) \equiv \int_{k_{\min}}^{1/\sqrt{A}} dk\, k P(k)/(2\pi) \propto \ln k/k_{\min}$. If each Fourier mode is statistically independent of every other Fourier mode, the density field will be the sum of many independent Fourier modes and will therefore be approximately Gaussian. However, for the sake of generality, we will not assume that δ is normally distributed. Instead, we assume only that p_A has a universal shape for all A. Since the mean of δ_A is zero for all A by definition, and since for any random variable θ the associated standard deviation obeys $\sigma_{a\theta} = a\sigma_\theta$, this allows us to write $p_A(\delta) = g(\delta/\sigma(A))/\sigma(A)$ for some general probability density function g. Differentiating f_A yields $n(N)$, the number of cities on Earth's surface with population $N = \bar{\rho}A$ per unit area per unit population:

$$n(N) = -\nu g(\nu)\frac{\rho}{N}\frac{d\ln\sigma}{dN} \propto \frac{1}{N^2}\frac{g(\nu)}{\ln(N_{\max}/N)}, \tag{A3}$$

where we have defined $\nu \equiv \delta_C/\sigma(N)$, the number of standard deviations associated with city formation. Note that, for $\nu \ll 1$, $g(\nu)$ is a slowly varying function of N for two reasons: the first derivative of g around $\nu = 0$ is small for small deviations from the mode, and ν is only a weak function of N. Thus, Eq. (A3) implies that the logarithmic slope $d\log n/d\log N$ tends to -2 in the limit of $N \ll N_{\max}$. This limit is empirically justified, since even the largest cities in the world contain only $\sim 10^{-3}$ of the world's population. Hence we arrive at Eq. (4).

Although our results are largely independent of the exact form of $p_A(\delta)$, let us briefly comment on its possible form. If $p_A(\delta)$ deviates from a Gaussian, this implies that different Fourier modes in human population density are correlated, a generic result of nonlinear interactions. Note, however, that $\delta \geqslant -1$ is strictly bounded from below, since human population density is always positive definite: $\rho \geqslant 0$. Hence, p_A cannot be exactly Gaussian. At some level, nonlinear interactions must come into play. If the population density fluctuations were typically small. $\delta \lesssim 1$, one might expect that a Gaussian distribution could be a good approximation; however, everyday experience tells us that population density fluctuations can be quite large. Indeed for New York City, $\delta \sim 300$. Hence, a theory of human population density growth must necessarily be nonlinear.

APPENDIX B: DERIVATION OF THE INVERSE-RANK FRIENDSHIP LAW

As a second application of our formalism, let us derive the average number of friends a person has in a given region. We again adopt a simple model, where we define a region to be a *community* if the population density exceeds some critical value $\delta \geqslant \delta_C$. This defines geographic equivalence classes on the inhabited regions, such that every person is a member of a community. Since real-world social networks are highly clustered and only a small fraction of people serve as connections between communities of friends [20], this assumption should be a good approximation for our purposes, since the more complicated topology of real-world friendship networks will mainly affect higher-order quantities that involve friends of friends and friends of friends of friends. Furthermore, we assume that the average number of friends D a given person has is asymptotically independent of the size of the community. This second assumption is essentially the assertion of the existence of the famous Dunbar's number [21,22], an upper limit on the number of people with whom a given person can sustain social relationships.

To compute the probability in the model, we consider two people A and B with N_{AB} people closer to A than B. If A is a member of a community with size $N_c \gg N_{AB}$, A and B are almost certainly friends. On the other hand, if A is in a community of size $N_c \ll N_{AB}$, it will be nearly impossible for A and B to be friends. There is thus a turnover scale at $\sim N_{AB}$ which dictates whether or not A and B will be friends; the probability is therefore determined by two independent events: the probability that A is in a community of size N greater than the turnover scale and the probability $p_f = D/N$ that A and B are friends given that A and B are in a community of size N, for large $N \gg D$. Since we know from the previous discussions that in such a model, the number density of communities scales asymptotically in proportion to $1/N^2$, and each community has N people, the probability p_c that a randomly chosen individual is in a community of size N scales as $\sim 1/N$. Hence,

$$\begin{aligned} p(N_{AB}) &= \int g(N,N_{AB})p_c(N)p_f(N)dN \\ &\propto \int_{N>N_{AB}} \frac{1}{N}\frac{D}{N}dN \propto \frac{1}{N_{AB}}, \end{aligned} \tag{B1}$$

where g has the properties that $0 \leqslant g < 1$, $g \approx 1$ for $N \gg N_{AB}$, and $g \ll 0$ for $N \ll N_{AB}$. The details of the function will depend on the geometry of the communities but do not concern us here as we are interested only in the scaling. We have thus derived the inverse-rank friendship law, previously proposed [23] to fit empirical data. We stress that our derivation is based entirely on theoretical considerations and therefore provides an explanation for the "physical" origin of the law.

APPENDIX C: TWO-POINT CORRELATION FUNCTION

In this appendix, we analytically compute the two-point correlation function [24] $\xi(x-y)=\langle\delta(x)\delta(y)\rangle$, which is the inverse Fourier transform of the power spectrum. The correlation function will play an important role in Appendix D. Physically, the correlation function measures the degree to which the existence of an overdensity or underdensity at some position $\mathbf{x}$ increases the likelihood that an overdensity or underdensity will be found at $\mathbf{y}$. Assuming circular symmetry, the inverse Fourier transform is a Hankel transform of order 0:

$$\xi(r)=\int_0^\infty \frac{k\,dk}{2\pi}\,P(k)J_0(kr), \tag{C1}$$

where J_0 is the first Bessel function. Taking $P(k)=P_0k^{-2}$ and a long-wavelength cutoff k_m gives us an integral that can be written in terms of special functions

$$\xi=\frac{P_0}{2\pi}\int_{k_m}^\infty dk\,\frac{J_0(kr)}{k}=\frac{P_0}{4\pi}G_{01}^{23}((k_mr/2)^2), \tag{C2}$$

where G is the Meijer G function. Defining a reduced area $a=k_m^2r^2/2$ and considering separations that are small compared to the system size (corresponding to the scale of continents) $a\ll 1$, we can expand

$$\xi(r)\approx\frac{P_0}{4\pi}\left[-\gamma+\frac{1}{2}\left(-\ln a+a-\frac{a^2}{8}+\frac{a^3}{108}\right)\right] \tag{C3}$$

where $\gamma\approx 5.7721$ is Euler's constant and we only neglect only terms $O((k_mr)^8)$. The second term guarantees that $\xi\gg 1$ for sufficiently small a and $\xi<0$ for $a\gtrsim 0.51$. Most importantly, we note that for r much smaller than the system size,

$$\xi(r)\to-\frac{P_0}{4\pi}\ln r. \tag{C4}$$

Since it is possible to invert a Fourier transform, any 2D model which predicts a correlation function that logarithmically diverges for small r must have a power spectrum of the form $P\propto k^{-2}$ for $k\gg k_m$.

APPENDIX D: EFFECTIVE FIELD THEORY

The basic program of effective field theory is the following: Given a statistical physics system in a fixed number of dimensions (in this case $D=2$), write down the Hamiltonian

$$H=\int d^2x\,\mathcal{H}(\delta,\nabla\delta,\nabla\nabla\delta,\dots), \tag{D1}$$

such that $\mathcal{H}$ contains all terms which are consistent with the symmetries of the system. By universality, the macroscopic properties of the system should then be reflected in the field theory. For a pedagogical introduction to this approach, see [12]. In our case, the symmetries are particularly constraining: we want the Hamiltonian to be invariant under scaling operations $x\to\lambda x$ in addition to translations and rotations of the Euclidean plane. Under a change of scale, the population density transforms like a scalar, so $\delta(x)\to\delta(\lambda^{-1}x)$ while $d^2x\to\lambda^2d^2x$ and $\nabla\delta\to\lambda^{-1}\nabla\delta$. Hence scale invariance requires that each term in $\mathcal{H}$ contain exactly two derivatives to cancel the λ^2 from the area element. Rotational symmetry then limits us to only one possible term $(\nabla\delta)^2$:

$$H=\frac{1}{2}\int d^2x\,(\nabla\delta)^2, \tag{D2}$$

which is simply a free scalar field in two dimensions. Addition of any interaction term to $\mathcal{H}$ of the form $V(\delta)$ is not allowed, as $d^2x\,V(\delta)$ would not transform correctly under a scale. Using standard field theory techniques, one can show that the correlation function $\xi(r)$ has the form of (C4); hence for wave vectors $k\gg k_m$ (corresponding to physical scales shorter than the system size), we must have that the power spectrum $P\propto k^{-2}$. This concludes our proof that a scalar random field in two spatial dimensions will have a power spectrum $P\propto k^{-2}$.

Let us make some further comments about (D2) that may be helpful to readers unfamiliar with effective field theory techniques. In particular, we can use (D2) to construct alternative theories that also will yield power spectra $P\propto k^{-2}$. For example, the generic Langevin equation, specialized to the case where the Hamiltonian is given by (D2), is just the famous diffusion equation with a noise term [12]:

$$\frac{\partial\delta}{\partial t}=-\lambda\nabla^2\delta+\eta, \tag{D3}$$

where δ is the fractional population overdensity and η is a fluctuating random variable. Notice that, while this model also involves a diffusion equation, it is conceptually distinct from Eq. (2). Here we think of population as physically diffusing in Cartesian space; in Eq. (2), the diffusion is not happening in Cartesian space but in Fourier space. Since (D3) was derived from a Langevin equation corresponding to the Hamiltonian (D2), its equilibrium properties must be described by (D2); hence it follows that $P(k)\propto k^{-2}$. Note, however, that the noisy diffusion model lacks parameters with dimensions of length to any power. Thus, our simple dimensional analysis argument presented after Eq. (3) still holds, and provides a simpler derivation of the power spectrum.

[1] L. M. Bettencourt, Science 340, 1438 (2013).
[2] G. K. Zipf, *Human Behavior and the Principle of Least Effort* (Addison-Wesley Press, Cambridge, MA, 1949).
[3] X. Gabaix, Q. J. Econ. 114, 739 (1999).
[4] B. Jiang and T. Jia, Int. J. Geographical Inf. Sc. 25, 1269 (2011).
[5] M. E. Newman, Contemp. Phys. 46, 323 (2005).
[6] M. Marsili and Y.-C. Zhang, Phys. Rev. Lett. 80, 2741 (1998).
[7] S. K. Baek, S. Bernhardsson, and P. Minnhagen, New J. Phys. 13, 043004 (2011).
[8] R. Gibrat, *Les Inégalités Economiques* (Recueil Sirey, Paris, France, 1931).
[9] J. M. Samuels, Rev. Econ. Stud. 32, 105 (1965).
[10] W. H. Press and P. Schechter, Astrophys. J. 187, 425 (1974).
[11] E. Vanmarcke, *Random Fields: Analysis and Synthesis* (World Scientific, Singapore, 2010).

[12] M. Kardar, *Statistical Physics of Fields* (Cambridge University Press, Cambridge, 2007).
[13] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, Astrophys. J. **304**, 15 (1986).
[14] E. Zambrano, A. Hernando, A. Fernández-Bariviera, R. Hernando, and A. Plastino, J. R. Soc. Interface **12**, 20150789 (2015).
[15] D. K. Ihrke and C. S. Faber, *Geographical Mobility: 2005 to 2010*, Technical Report (U.S. Department of Commerce, Washington, DC, 2012).
[16] NASA Socioeconomic Data and Applications Center, *Gridded Population of the World*, Technical Report (Center for International Earth Science Information Network, Centro Internacional de Agricultura Tropical, Palisades, NY, 2010),http://sedac.ciesin.columbia.edu/data/collection/gpw-v3.
[17] R. K. Sheth and G. Tormen, Mon Not. R. Astron. Soc. **308**, 119 (1999).
[18] J. R. Bond, S. Cole, G. Efstathiou, and N. Kaiser, Astrophys. J. **379**, 440 (1991).
[19] A. Loeb and S. R. Furlanetto, in *The First Galaxies in the Universe*, edited by A. Loeb and S. R. Furlanetto (Princeton University Press, Princeton, NJ, 2013).
[20] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[21] R. Dunbar, J. Human Evolution **22**, 469 (1992).
[22] B. Gonçalves, N. Perra, and A. Vespignani, PloS One **6**, e22656 (2011).
[23] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, and A. Tomkins, Proc. Natl. Acad. Sci. U.S.A. **102**, 11623 (2005).
[24] D. M. Smith, Jr., Commun. Astron. Public J. **15**, 31 (2014).