PHYSICAL REVIEW E **96**, 050301(R) (2017)

Optimal paths on the road network as directed polymers

A. P. Solon,[1] G. Bunin,[2] S. Chu,[1] and M. Kardar[1]
[1]*Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[2]*Department of Physics, Technion, Haifa 32000, Israel*
(Received 1 June 2017; published 1 November 2017)

We analyze the statistics of the shortest and fastest paths on the road network between randomly sampled end points. We find that, to a good approximation, the optimal paths can be described as directed polymers in a disordered medium, which belong to the Kardar-Parisi-Zhang universality class of interface roughening. Comparing the scaling behavior of our data with simulations of directed polymers and previous theoretical results, we are able to point out the few characteristics of the road network that are relevant to the large-scale statistics of optimal paths. Indeed, we show that the local structure is akin to a disordered environment with a power-law distribution which become less important at large scales where long-ranged correlations in the network control the scaling behavior of the optimal paths.

DOI: 10.1103/PhysRevE.96.050301

Complex networks of nodes and links can be used to model a wide array of systems. Examples range from biological networks such as those formed by neurons and synapses in the brain or chemical reactions inside a cell, to social or transportation networks and the World Wide Web. Their topology in the abstract space of edges and vertices has been much studied, allowing one to identify widespread properties such as "small-world" effects, scale-free connectivity, and a high degree of clustering, which can be captured by simple physical models [1–5]. Comparatively, less is understood about the spatial organization of complex networks embedded in a Euclidean space, a very active subject of research (see Ref. [6] for a review). The effect of geometry becomes especially relevant when the network is strongly constrained by the environment or when the "cost" to maintain edges increases significantly with their length (e.g., rivers [7], railways [8], or vascular networks [9]). The spatial structure of streets is another example that has been particularly studied to gain insight into the structure of cities and their development [10–12].

Much information about the geometry of a network can be obtained by studying the shortest paths between the nodes of the network. In many cases, it is also a problem of practical importance to characterize the paths that optimize a given cost function. For example, in transportation networks, one would like to understand the properties of the paths that minimize the travel time, the distance, or the monetary cost to travel between two points. An obvious application is in the development of efficient global positioning system (GPS) routing algorithms which could use prior information on optimal paths to perform better [13]. The shortest paths between two generating nodes on the power grid are also important to predict the overloading of electric lines [14]. Understanding the properties of these optimal paths appears challenging since they are expected to depend strongly on the geometry of the network which can be shaped by various factors, from natural obstacles to historical development or differences in policy.

The theory of directed polymers [15] tackles a related problem. It is concerned with the statistics of a chain stretched between two points that minimizes its energy in a random environment modeled by a fluctuating potential. The optimal configuration is then found as a trade-off between the line tension of the chain and the energy imparted by the random environment. A wealth of theoretical results is available for directed polymers which belong to the Kardar-Parisi-Zhang (KPZ) [16] universality class describing the roughening of growing interfaces, maybe the most studied class of systems in nonequilibrium statistical physics.

In this Rapid Communication, we study the statistics of optimal (shortest and fastest) paths on the road network in light of known results for directed polymers in a random medium (DPRM). Gathering large data sets of millions of paths on three continents, we compute the probability distribution of path length and travel time as a function of the distance between the end points. By comparison with a DPRM model, we show that most details of the structure of the road network are not relevant to the statistics on larger scales. The local environment can be modeled by a power-law distributed noise with, remarkably, a universal decay exponent. Furthermore, we show that long-range correlations in the environment, on the scale of hundreds of kilometers, affect the scaling exponents and are thus relevant to the statistics of optimal paths. The transverse wandering of the paths is also found to be consistent with our modeling as a directed polymer.

Let us first introduce more precisely the two-dimensional (2D) DPRM problem and summarize the results relevant to our study. A directed polymer is a chain pinned at its end points that is sufficiently stretched to prevent overhangs and thus can be described by a scalar height function $h(x)$, with x a coordinate along the axis going between the end points and h the distance to the axis. The energy of a configuration of the chain is given by

$$E[h(x)] = \int_0^d \left[\frac{\gamma}{2} \left(\frac{dh}{dx} \right)^2 + V(x,h) \right] dx, \qquad (1)$$

where d is the distance between the end points, γ is related to the line tension of the chain, and V is a random potential modeling a disordered environment. The free energy of such configurations follows the KPZ equation [16] that gives rise to scale invariance. (Because of this mapping, x is traditionally denoted t as a time direction, but we stick here to the spatial notation to avoid confusion with travel times.) In the zero-temperature limit that is relevant to our problem,

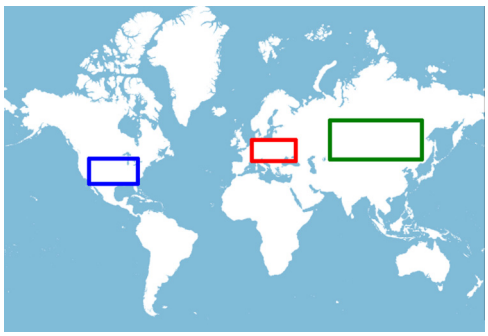

FIG. 1. The location of the three regions considered. For simplicity and efficiency of our algorithm, they are chosen to be large rectangular areas (in latitude-longitude coordinates) without sea or ocean.

the free energy is simply the energy of the optimal path $E[h^*]$. Two exponents govern the scaling of the energy fluctuations $\langle (E - \langle E \rangle)^2 \rangle \sim d^{2\beta}$ (where the brackets denote an average over realizations of the disorder V) and the transverse wandering of the optimal chains, $\langle h^*(x)^2 \rangle \sim x^{2\zeta}$ [17]. In 2D, if V has only short-ranged correlations, the exponents $\beta = 1/3$ and $\zeta = 2/3$ are known exactly [18]. More recently, it has been shown that the full distribution of E is actually universal, converging at large d to Tracy-Widom (TW) distributions of random matrix theory [19,20]. On the contrary, long-range correlated disorder leads to larger scaling exponents and different energy distributions [21–25].

In light of these theoretical results, we now analyze the statistics of two types of optimal paths (the shortest and the fastest) on the road network. We compute the paths using the Open Source Routing Machine (OSRM) [26] operating on OpenStreetMap data, a collaborative effort to provide an open-source map of the world. The fastest paths are determined using the default configuration of OSRM which takes into account speed limitations for cars and road types but no information on traffic. We gather six data sets for the two types of optimal paths in the three regions indicated in Fig. 1, sampling the end points of the paths uniformly on the network.

In Fig. 2, we show examples of optimal paths drawn from an arbitrary center point (near Munich, Germany) to uniformly sampled points at a 300 km distance. Both sets of optimal paths

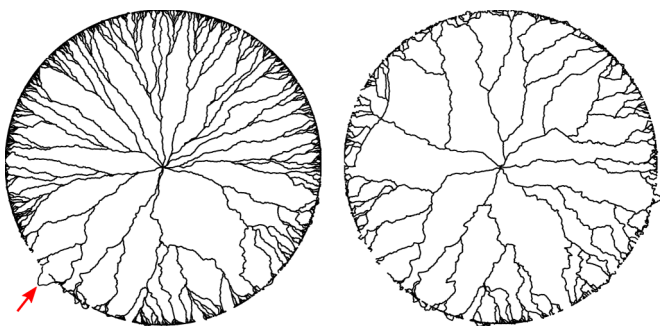

FIG. 2. Shortest (left) and fastest (right) paths from a center point (near Munich, Germany) to 10^4 randomly chosen points at a distance of 300 km. The arrow points to the most prominent overhang in the paths.

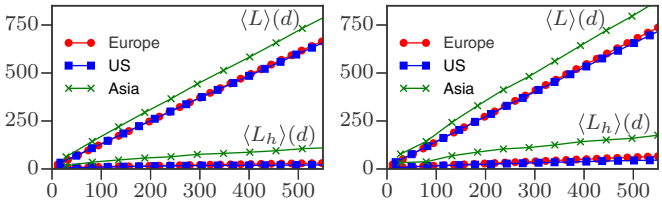

FIG. 3. Average length of the optimal paths $\langle L \rangle$ and length of overhangs $\langle L_h \rangle$ as a function of the distance d between the end points. All lengths are measured in km. Left: Shortest paths. Right: Fastest paths. 10^6 points for each curve. Lines are guides to the eye.

display a fractal branching pattern strongly resembling what is observed in directed polymer models [17]. However, these routes are not perfectly directed. This is especially visible near the end points where the local structure of the road network may impose overhangs (the most prominent is indicated by a red arrow in Fig. 2). Nevertheless, overhangs are mostly avoided by the optimal paths on the rest of the trajectory, as quantified in Fig. 3 where we plot the average length of the paths $\langle L \rangle(d)$ and the part $\langle L_h \rangle(d)$ corresponding to overhangs (see the Supplemental Material [27] for a precise definition of overhangs). The average path length $\langle L \rangle$ is found to increase linearly with d at large distances while $\langle L_h \rangle$ increases slower. Overhangs thus become less relevant at larger distances for which we expect a better comparison between road paths and directed polymers. In the following, we divide accordingly our study between short paths that are strongly constrained by the network, and longer paths which result from optimization.

We first look in Fig. 4 at the distribution of the length L of the shortest paths (respectively the travel time T on the fastest paths) between points at a short distance $d = 1$ km. We are interested in L and T as the quantities that are minimized and thus, in our interpretation, akin to the energy of a directed chain. The distributions display clear power-law tails at large L and T over more than three orders of magnitude. The tails correspond to situations where the path has to go around an obstacle to reach a nearby point, e.g., reach the next bridge to cross a river. They thus characterize the overhangs described previously. Most remarkably, the decay exponent $P(L) \sim L^{-\alpha}$ [and $P(T) \sim T^{-\alpha}$] seems to be universal across continents with $\alpha \approx 3$ (the best-fit coefficients for the six curves are all found within $[2.83; 3.10]$). This appears surprising since we expect the paths at small d to reflect the local structure of the road network which is *a priori* very different in the three regions considered. Although we lack an explanation for the value of the exponent, it can be compared to exponents derived for the same distribution in different environments. The shortest path between nearby points on the backbone of a percolation cluster has been numerically found to exhibit the same $\alpha = 3$ [28] at small distances, while for self-avoiding random walks, the probability of forming a loop of length ℓ in a 2D chain scales as $\ell^{-\alpha}$ with (exact) exponent $\alpha = 2.68$ [29]. The latter superficially resembles the configuration of a road between nearby points that loops around to avoid intervening obstacles, while not intersecting other roads arriving and/or departing the two points.

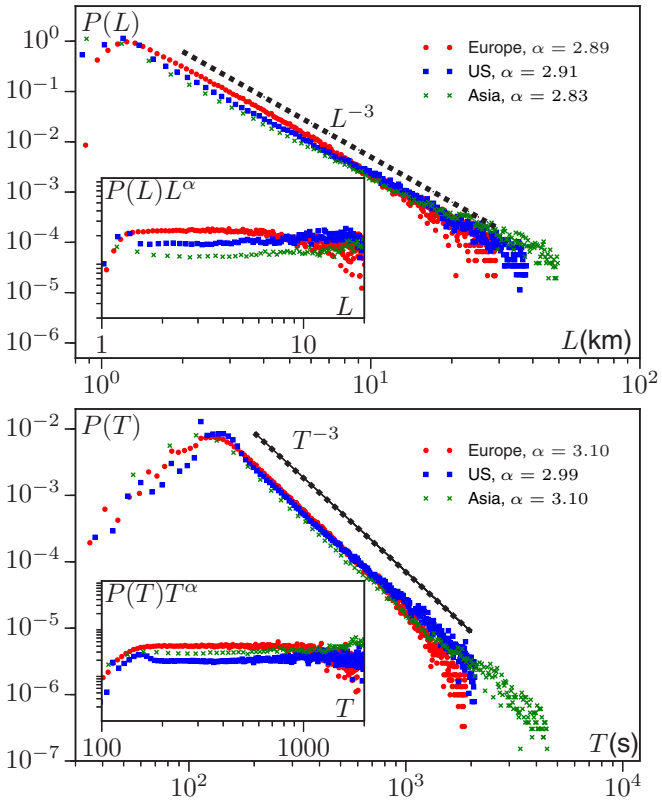

FIG. 4. Optimal paths between points at a distance $d = 1$ km. Top: Probability distribution of the length L of the shortest paths. Bottom: Probability distribution of the travel time T of the fastest paths. $N = 5 \times 10^5$ paths for each curve. Insets: Scaled plots with the best-fit exponent α indicated in the legend.

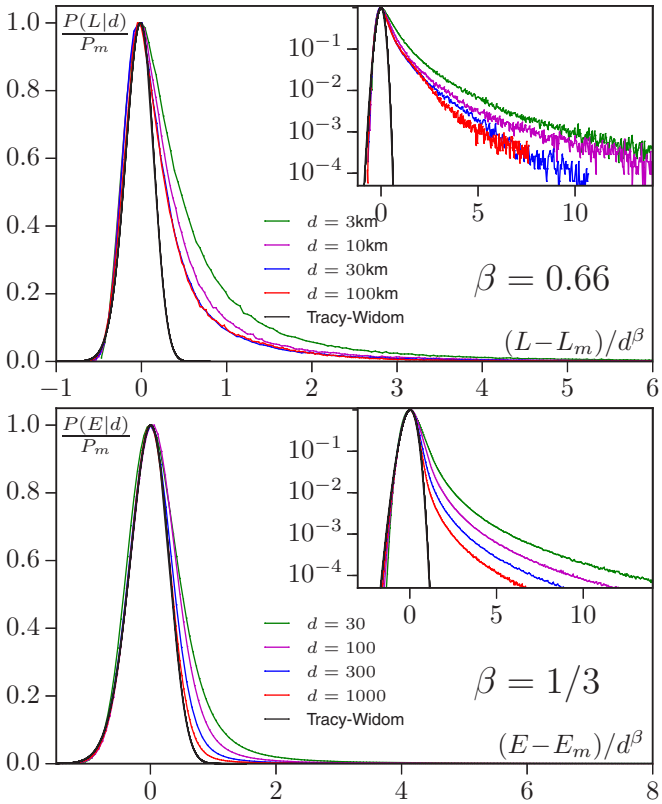

FIG. 5. Top: Probability distribution of the length L of the shortest paths in Europe rescaled with $\beta = 0.66$. 5×10^5 paths for each curve. Bottom: Probability distribution of the energy for the DPRM model with power-law noise rescaled with $\beta = 1/3$. 10^7 paths for each curve. Lattice of size 10^7 in the transverse direction with periodic boundary conditions. The insets show the same data with a logarithmic y axis. P_m denotes the maximum of the distribution, found at a value $L = L_m$ or $E = E_m$.

Because of the fat tails in the distributions of Fig. 4, the variance of L and T is not defined. We thus cannot estimate the exponent β characterizing energy fluctuations by simply looking at the scaling of $\langle L^2\rangle_c(d)$ and $\langle T^2\rangle_c(d)$, and need to look instead at the full probability distributions $P(L|d)$ and $P(T|d)$ for increasing distance d between the end points. To compare these distributions, we superimpose their maxima and rescale their width by a factor d^β where β is adjusted so that the distributions converge at large d. The results are shown in Fig. 5 (top) for the shortest paths in Europe and in the Supplemental Material [27] for the five other data sets, which show similar behavior. We find that the exponent β can be adjusted such that the left tail of the distribution converges rapidly to a limit distribution well fitted by the Tracy-Widom (TW) distribution expected for directed polymers. On the contrary, the right tail converges slower and remains heavy at the largest d attainable (larger d, comparable to the total size of the region, show strong finite-size effects). It is thus not clear if the right tail also converges to TW behavior or to a different distribution, as was observed numerically for a directed polymer model with long-ranged correlations in the environment [25].

For comparison, we simulated a well-established DPRM model on a square lattice with random independent energies on each site [17,18,30]. The paths are directed in the diagonal of the lattice, parametrized by d. As before, the distance (in number of sites) from the diagonal is denoted by h. The energy of the optimal path is computed recursively as

$$E(d,h) = \min\{E(d-1,h), E(d-1,h-1)\} + \eta(d,h). \quad (2)$$

After d iterations, $E(d,h)$ is then the energy of the optimal path between the point (d,h) and the line $d = 0$. As opposed to previous studies that considered Gaussian noise, we draw the noise $\eta(d,h)$ in a power-law distribution $P(\eta) = 2\eta^{-3}$ with $\eta \in [1;\infty[$ to match qualitatively the short-scale distributions in Fig. 4. We then analyze the results as in the experimental case: We shift the energy distributions $P(E|d)$ to superimpose their maxima and rescale their width by d^β (Fig. 5, bottom). We observe that, as with Gaussian noise [30], the distribution converges to a TW distribution with the KPZ exponent $\beta = 1/3$. Indeed, only a fat tail in the noise at *negative* energy, $P(\eta) \sim \eta^{-a}$, as $\eta \to -\infty$ would change the scaling exponents [31,32]. Interestingly, the convergence when increasing d happens in a similar manner in the model and the experimental data, with the right tails converging much slower. This also lends credit to our measure of β as the exponent rescaling the left tail of the distributions.

One salient difference remains between the paths on the road and the directed polymer model: The measured β

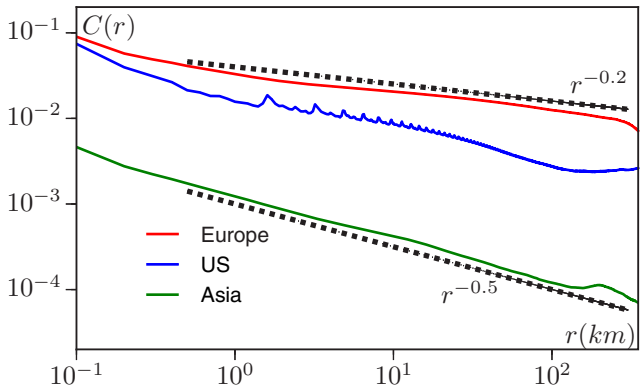

FIG. 6. Autocorrelation function of the road density as defined in the text. The oscillations in the curve for the U.S. are not an artifact. The peaks are located every mile (with subpeaks at half miles) and correspond to large regions (up to 60 miles) of gridlike road network.

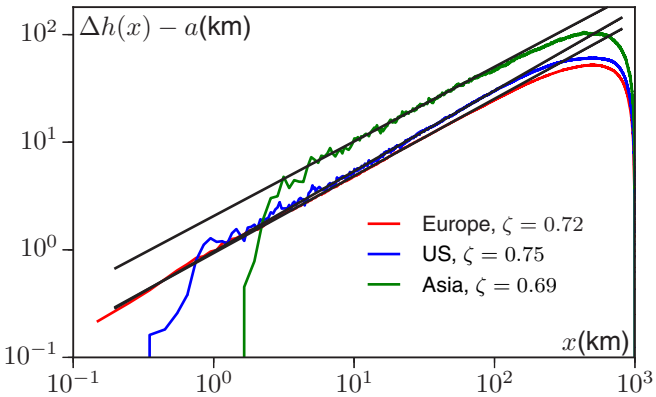

FIG. 7. Transverse wandering for the shortest paths as a function of the coordinate x on the axis between the end points. Average over 5×10^5 paths between points at distance $d = 1000$ km.

exponents are found between 0.58 and 0.9 (with an error estimated around 15%) and are thus much larger than $\beta = 1/3$ in the KPZ universality class. We now argue that this can be explained by the presence of long-range correlations in the road network. To show this, we first discretize the full map of each region in squares of size 100 m × 100 m and assign the value $\rho(\mathbf{r}) = 1$ if a road is found inside the square and 0 otherwise. We then compute the correlation function $C(r) = \langle \rho(\mathbf{r})\rho(\mathbf{r}+\mathbf{x})\rangle - \langle \rho(\mathbf{x})\rangle^2$ where the average is taken over $\mathbf{x}$ and orientations of $\mathbf{r}$. As shown in Fig. 6, $C(r)$ decreases slowly [slower than $C(r) \sim r^{-0.5}$], remaining non-negligible on the scale of hundreds of kilometers. These long-range correlations reflect the fact that the road network is shaped by many factors acting at every scale, from different administrative divisions to natural obstacles. They were also shown to be important in modeling the development of cities [33], obviously related to that of the road network.

For DPRM, such a slow decay of correlations was proven to be relevant to the large-scale behavior, both in numerical simulations [23,25,34] and analytic calculations [21,24,35,36]. For Gaussian noise with isotropic correlations decaying as a power law with exponents between -0.5 and -0.2 (as measured for the road density correlations in Fig. 6), β was measured between 0.5 and 0.7 [23]. Given experimental uncertainties, these values are in relatively good agreement with our measurements for the road network. Long-range correlations are thus likely to be at the origin of the large exponents observed.

Finally, we look at the wandering of the optimal paths in the transverse direction. The routing algorithm returns a list of points along each path (in average every 50 m) that we use to construct the function $h(x)$, the distance to the end-to-end axis parametrized by x. We do so by discretizing x in bins of size $dx = 100$ m and averaging points falling in the same bin. This discards any overhang and thus produces a directed path approximating the real path. The leading behavior is expected to be scale invariant, $\Delta h(x) = \sqrt{\langle h^2(x)\rangle} \sim x^{\zeta}$. However, because of overhangs near the end points, $h(0) \neq 0$, so that $\Delta h(0) \neq 0$, inducing large corrections to the putative scaling. Thus, as a first approximation, we estimate the ζ exponent by fitting $\Delta h(x) = a + bx^{\zeta}$ with free parameters a, b, and ζ. The resulting functions $\Delta h(x) - a$ show scaling behavior over two orders of magnitude with exponents $\zeta \in [0.69; 0.75]$ (see Fig. 7). Once again these values are larger than the KPZ exponent $\zeta = 2/3$, in qualitative agreement with the presence of long-range correlations that are expected to increase the value of ζ. For comparison, isotropic long-range correlations with a decay exponent in the range of Fig. 6 give $\zeta \in [0.75; 0.85]$ [23] while correlations only in the transverse direction yield $\zeta \in [0.67; 0.72]$ [25].

To conclude, we have shown that optimal paths on the road network can be modeled as directed polymers in a random medium. In doing so, we replaced the complex road structure by a homogeneous noise featuring only the relevant properties to account for the observed statistics of optimal paths. We find two important characteristics. At short distances, the structure of the network induces a scale-free distribution of path length with a universal decay exponent, a remarkable experimental fact that remains to be explained. This is accounted for by a power-law distributed noise in our DPRM model. At larger scales, the local structure becomes less relevant. The scaling of the path length or travel time and the transverse wandering of optimal paths are then governed by long-range correlations that we show to be present in the network. Although these long-range correlations are nonuniversal, they show similar behaviors in the different regions of the world considered here, leading to similar distributions at large scales. Directed polymers and associated theoretical results thus provide useful tools to understand the statistics of optimal paths on a complex network. It would be interesting in the future to see if this approach can be extended to other transportation networks or different environments, for example, to the study of shortest paths on a critical percolation cluster [37], which have important practical applications.

A.P.S. thanks J. B. Deris for insightful discussions and the Gordon and Betty Moore Foundation for funding through a PLS fellowship. S.C. and M.K. acknowledge funding from the NSF Grant No. DMR-1708280.

[1] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[2] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[3] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[4] E. Ravasz and A.-L. Barabási, Phys. Rev. E **67**, 026112 (2003).
[5] M. E. Newman, SIAM Rev. **45**, 167 (2003).
[6] M. Barthélemy, Phys. Rep. **499**, 1 (2011).
[7] P. S. Dodds and D. H. Rothman, Phys. Rev. E **63**, 016115 (2000).
[8] P. Sen, S. Dasgupta, A. Chatterjee, P. A. Sreeram, G. Mukherjee, and S. S. Manna, Phys. Rev. E **67**, 036106 (2003).
[9] D. Hunt and V. M. Savage, Phys. Rev. E **93**, 062305 (2016).
[10] A. Cardillo, S. Scellato, V. Latora, and S. Porta, Phys. Rev. E **73**, 066107 (2006).
[11] S. Lämmer, B. Gehlsen, and D. Helbing, Physica A **363**, 89 (2006).
[12] M. Barthélemy and A. Flammini, Phys. Rev. Lett. **100**, 138702 (2008).
[13] R. Geisberger, P. Sanders, D. Schultes, and D. Delling, in *Experimental Algorithms: 7th International Workshop*, WEA 2008 Provincetown, MA, USA, May 30–June 1, 2008, edited by C. C. McGeoch (Springer, Berlin, 2008), pp. 319–333.
[14] C. Schiel, P. G. Lind, and P. Maass, Sci. Rep. **7**, 11562 (2017).
[15] T. Halpin-Healy and Y.-C. Zhang, Phys. Rep. **254**, 215 (1995).
[16] M. Kardar, G. Parisi, and Y.-C. Zhang, Phys. Rev. Lett. **56**, 889 (1986).
[17] M. Kardar and Y.-C. Zhang, Phys. Rev. Lett. **58**, 2087 (1987).
[18] D. A. Huse, C. L. Henley, and D. S. Fisher, Phys. Rev. Lett. **55**, 2924 (1985).
[19] M. Prähofer and H. Spohn, Phys. Rev. Lett. **84**, 4882 (2000).
[20] K. A. Takeuchi, M. Sano, T. Sasamoto, and H. Spohn, Sci. Rep. **1**, 34 (2011).
[21] E. Medina, T. Hwa, M. Kardar, and Y.-C. Zhang, Phys. Rev. A **39**, 3053 (1989).
[22] C.-K. Peng, S. Havlin, M. Schwartz, and H. E. Stanley, Phys. Rev. A **44**, R2239 (1991).
[23] R. Schorr and H. Rieger, Eur. Phys. J. B **33**, 347 (2003).
[24] T. Kloss, L. Canet, B. Delamotte, and N. Wschebor, Phys. Rev. E **89**, 022108 (2014).
[25] S. Chu and M. Kardar, Phys. Rev. E **94**, 010101(R) (2016).
[26] D. Luxen and C. Vetter, in *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (ACM, New York, 2011), pp. 513–516.
[27] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.96.050301 for complementary figures and technical details.
[28] M. Porto, S. Havlin, H. E. Roman, and A. Bunde, Phys. Rev. E **58**, R5205 (1998).
[29] B. Duplantier, Phys. Rev. B **35**, 5290 (1987).
[30] J. M. Kim, M. A. Moore, and A. J. Bray, Phys. Rev. A **44**, 2345 (1991).
[31] T. Gueudre, P. Le Doussal, J.-P. Bouchaud, and A. Rosso, Phys. Rev. E **91**, 062110 (2015).
[32] N.-N. Pang and T. Halpin-Healy, Phys. Rev. E **47**, R784(R) (1993).
[33] H. A. Makse, J. S. Andrade, M. Batty, S. Havlin, and H. E. Stanley, Phys. Rev. E **58**, 7054 (1998).
[34] T. Song and H. Xia, J. Stat. Mech.: Theory Exp (2016) 113206.
[35] E. Katzav and M. Schwartz, Phys. Rev. E **70**, 011601 (2004).
[36] A. A. Fedorenko, Phys. Rev. B **77**, 094203 (2008).
[37] M. Barma and P. Ray, Phys. Rev. B **34**, 3403 (1986).