Scale invariance in road networks

Vamsi Kalapala,[1,*] Vishal Sanwalani,[1,†] Aaron Clauset,[1,‡] and Cristopher Moore[1,2,§]

[1]*Department of Computer Science, University of New Mexico, Albuquerque, New Mexico 87131, USA*
[2]*Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico 87131, USA*

We study the topological and geographic structure of the national road networks of the United States, England, and Denmark. By transforming these networks into their *dual* representation, where roads are vertices and an edge connects two vertices if the corresponding roads ever intersect, we show that they exhibit both topological and geographic scale invariance. That is, we show that for sufficiently large geographic areas, the dual degree distribution follows a power law with exponent $2.2 \leqslant \alpha \leqslant 2.4$, and that journeys, regardless of their length, have a largely identical structure. To explain these properties, we introduce and analyze a simple fractal model of road placement that reproduces the observed structure, and suggests a testable connection between the scaling exponent α and the fractal dimensions governing the placement of roads and intersections.

PACS number(s): 89.75.Da, 47.53.+n, 89.40.Bb, 89.75.Hc

I. INTRODUCTION

Complex networks has received much attention from the physics community and beyond in the recent past [1–3]. This interest has primarily sprung from the near ubiquity of networks in both the natural and manmade world. Canonical examples of complex networks include the Internet [4], the World Wide Web [5], social contacts [6], scientific citations [7,8], and gene and protein interactions [9,10]. Most of these studies have focused on topological quantities like the degree distribution, diameter, and clustering coefficient. More often than not, it has been found that networks exhibit a degree distribution in which the fraction of vertices with degree k has the form of a power law, $P(k) \sim k^{-\alpha}$, where $2 < \alpha < 3$.

While virtual networks like the World Wide Web, or interaction networks like that of proteins, may be considered purely in terms of their topology, physical networks have additional geographic properties. In particular, creating and maintaining edges presumably requires physical resources proportional to their length, and the physical length of a path between two vertices may be rather different from its topological length (i.e., the number of edges along it). In some cases, the interaction of a network's topology with its underlying geography has been studied previously through models of evolving networks or optimizing resource costs [11–13].

Here, we focus on the presence of hierarchy and scale invariance in physical networks as illustrated by the nationwide road networks of the United States, England, and Denmark. To reveal their topological organization, we employ the *dual* model of the road network, in which a vertex represents a single road of a given name, and two vertices are joined if their corresponding roads ever intersect. This should not be confused with the dual of a planar graph, in which faces become vertices and vice versa. This graph

transformation has been used previously to study the topological structure of urban roads [14–17].

By representing the road network in this manner, we are able to show empirically that the degree distribution has a heavy tail, and is well characterized by a power law with an exponent $2.2 \leqslant \alpha \leqslant 2.4$. Rosvall *et al.* showed that urban networks also have heavy tails in the dual degree distribution, although not unequivocally with a power-law form [16]. Additionally, we find the structure of journeys on the physical network is scale invariant, i.e., the structure of a journey is similar regardless of its scale. To explain these properties, we introduce and analyze a simple fractal model for the hierarchical placement of roads on the unit square. We show that the recursive nature of this model generates the scale invariant journey structure, and suggests a simple relationship between the scaling exponent of the dual degree distribution α and the fractal dimensions governing the placement of roads and intersections.

II. PRIMAL AND DUAL MODELS

The natural representation of a road network is a collection of road segments, in which each segment terminates at an intersection; this is called the *primal* representation. However, this representation gives us little opportunity to consider scale-free properties or heavy-tailed degree distributions: almost all vertices have degree 4, and the average degree of a planar network is at most 6, as the maximum number of edges is $3n-6$. However, this representation violates the intuitive notion that an intersection is where two roads cross, not where four roads begin. Nor does it well represent the way we tell each other how to navigate the road network [18], e.g., "stay on Main Street for 10.3 miles, ignoring all cross streets, until you reach Baker Street, then turn left." If we use the dual representation, however, such a set of directions reduces to a short path through the network where each transition from one road to another corresponds to a single step.

In order to transform the road network into its dual representation, we must define which road segments naturally

*Electronic address: vamsi@cs.unm.edu
†Electronic address: vishal@cs.unm.edu
‡Corresponding author. Electronic address: aaron@cs.unm.edu
§Electronic address: moore@cs.unm.edu

belong together. In previous studies of road networks, segments have been grouped by their street name [15,16], line of sight by a driver [19], or by using a threshold on the angle of incidence of segments at an intersection [14]. Here, we use the method of taking a single road to be the collection of road segments that bear the same street name.

III. SAMPLING METHODOLOGY

We sampled the national road networks of the United States, England, and Denmark by querying a commercial service, provided by Mapquest.com. This service provides driving directions, i.e., a path through the dual graph, when given source and destination addresses. If only partial information is provided, e.g., the postal code, the service defaults to a unique address near that region's center. The directions returned are a list of road names, the respective distances a driver should travel on each, and instructions as to how to get from one road to another, e.g., "turn left onto" or "continue on."

We sampled each road network by taking the union of paths between a large number of uniformly random pairs of source and destination postal codes. For the United States, we sampled until every postal code was present in the network (roughly 200 000 trials), while for England and Denmark, we sampled beyond this limit (for about 35 000 total, each). We repeated our analysis for a random fraction 0.25, 0.50, and 0.75 of each sample and found that our observations are not sensitive to the number of trials.

Notably, our sampled networks are biased according to population distribution, as postal codes in each country are distributed roughly according to population. On the other hand, by focusing only on travel between postal regions, we restrict ourselves to studying the structure of long journeys. Naturally, we expect short-range travel to represent the majority of real journeys, e.g., trips to the office, the grocery store, etc. Finally, while most details of the algorithm that Mapquest uses to generate its driving directions are concealed on account of it being proprietary, we note that any algorithm that minimizes travel time, as opposed to geographical distance, will create a bias toward traveling on major roads and highways.

IV. JOURNEY STRUCTURE

Intuitively, a road network is composed of a hierarchy of roads with different importance. For instance, a road atlas classifies roads, according to their speed limit and capacity, into minor and major local streets, regional roads, and finally highways. Assuming that a driver wishes to reach her destination as quickly as possible, we may model the structure of an arbitrary journey as follows. Our driver begins at the local street where her point of origin is located, and moves to progressively larger and faster roads, i.e., she moves up the hierarchy, until she reaches the fastest single road between her source and destination. On this road, she covers as much distance as possible, and then descends to progressively smaller roads until she reaches the local street of her destination.
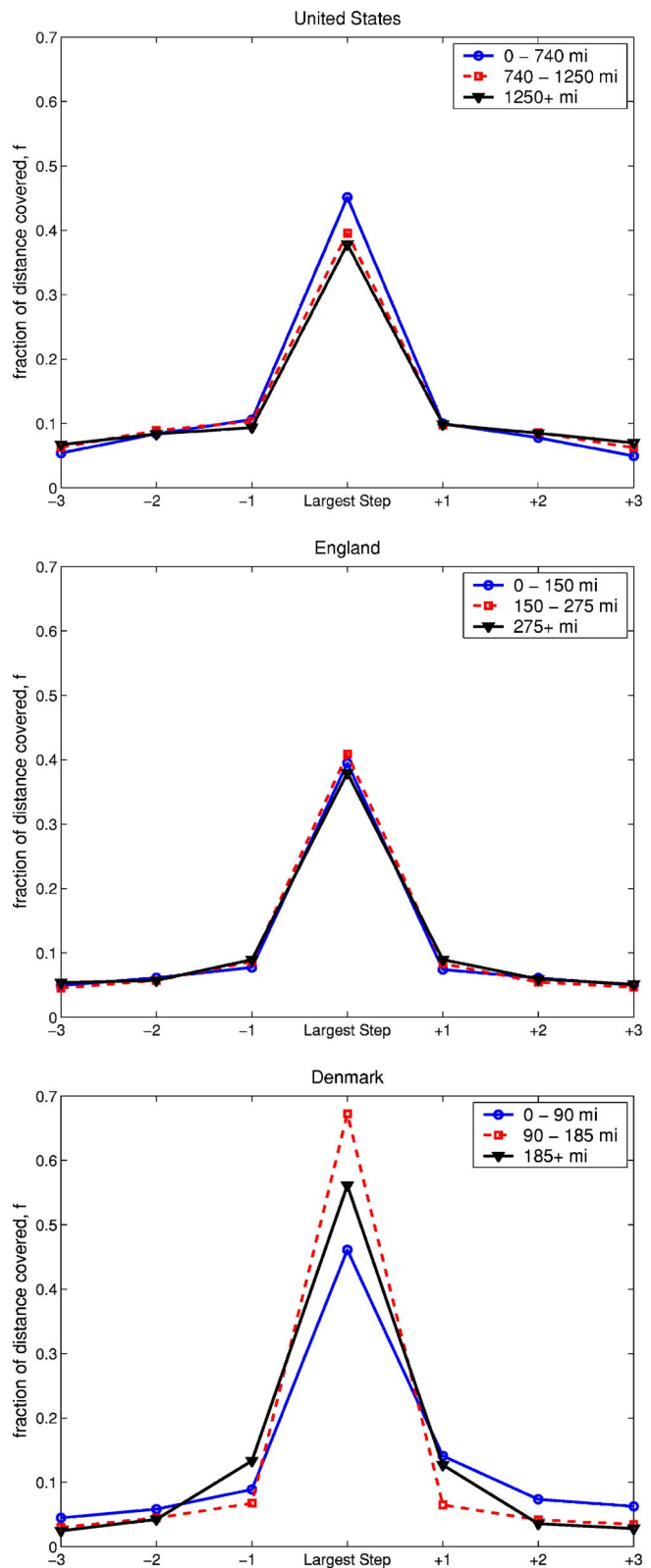

FIG. 1. (Color online) The average journey profiles for the United States, England, and Denmark. Profiles are defined to be the largest step (centered) in the journey, flanked by the three largest preceding and subsequent steps (in order of appearance) in the path.

TABLE I. The average fraction $A_j = s_j/\ell$ of the total length covered by each of the five largest steps with standard deviations.

Network	Distance	1st largest	2nd largest	3rd largest	4th largest	5th largest	Sum
	0–750	0.460±0.179	0.212±0.081	0.118±0.056	0.071±0.040	0.045±0.030	0.905
United States	750–1250 mi	0.397±0.149	0.208±0.072	0.129±0.048	0.084±0.038	0.055±0.030	0.873
	1250+	0.382±0.153	0.201±0.061	0.132±0.048	0.083±0.036	0.058±0.030	0.856
	0–150	0.396±0.148	0.088±0.042	0.061±0.030	0.061±0.030	0.046±0.024	0.739
England	150–275 mi	0.410±0.205	0.169±0.087	0.084±0.043	0.057±0.031	0.042±0.024	0.761
	275+	0.390±0.014	0.158±0.062	0.097±0.033	0.063±0.021	0.044±0.018	0.752
	0–90	0.472±0.183	0.199±0.085	0.098±0.040	0.070±0.029	0.052±0.027	0.891
Denmark	90–185 mi	0.672±0.179	0.139±0.088	0.066±0.051	0.037±0.026	0.025±0.018	0.939
	185+	0.564±0.100	0.253±0.118	0.064±0.040	0.034±0.021	0.023±0.014	0.937

Thus, we expect that the largest steps of a journey will cover a significant fraction of the total distance, and that the length of a step will increase as a driver moves up the hierarchy in the beginning of the journey, and decrease as she descends it at the journey's end. Empirically, we find that this assumption reflects the structure of journeys through our sampled networks. For the purposes of comparison, we classify journeys into three roughly equally populated groups based on their length: short, medium, and long.

To more precisely compare the journey structure between trips of different lengths, we define a journey's *profile* in the following way. We take the largest step of the journey, in terms of distance traveled, the three largest steps (in order of appearance) that precede it, and the three largest steps (again, in order of appearance) that follow it. Thus we ignore the many small steps that are scattered throughout the journey, e.g., taking a highway ramp to merge onto a national highway. While this definition of a journey profile is somewhat arbitrary, it allows us to focus on the journey's large-scale structure.

Figure 1 illustrates the average profile for journeys on each of the three national road networks for short, medium, and long journeys. The unimodal shape of these profiles clearly supports the hierarchical model we describe above. Additionally, their approximate collapse across journeys of different lengths indicates that the structure of the journey profile is invariant with respect to the scale of the journey. We omit error bars in Fig. 1 for visual conciseness, but note that they are less than ±0.1, on average. The slight asymmetry in Denmark's short-journey profile reflects the presence of one-way roads of different lengths.

In Table I, we show the fraction of the total distance covered by the five largest steps of these journeys, where s_j is the jth largest step. These steps alone account for about 85% of the total length of the journey; the largest step typically covers about 46% of the entire distance, the second largest covers 19%, the third largest covers 10%, etc. Moreover, for each j from 1 to 5, the fraction of the journey covered by the jth largest step appears to be roughly constant. This suggests a simple linear relationship of the form

$$s_j = A_j \ell, \qquad (1)$$

where s_j is the jth largest step, ℓ is the total path length, and A_j is some constant. Figure 2 shows the average step size for

each of the five largest steps against the total path length for each of our three networks. We fit our data to a power law with the form $s_j = A_j \ell^{\alpha_j}$, bootstrapped via least squares (we ignore the longest journeys, since we expect finite-size effects to appear as ℓ approaches the diameter of the country). We observe that this power law fits the data reasonably well, with average r^2 values of 0.97, 0.99, and 0.99, respectively; moreover, averaging across all such models, we have $\alpha_j = 1.0 \pm 0.1$, suggesting that the linear form of (1) is accurate.

V. DEGREE DISTRIBUTION

Other studies of road networks have found that the degree distribution of the dual graph, i.e., the number of intersections in which a single road is involved, is heavy-tailed, although not necessarily a power law [15,16]. We find similarly heavy-tailed distributions at the national level (Fig. 3), but with apparent finite-size cutoffs related to respective geographic scales. For small countries like England and Denmark, and for cities as in [16], the case for scale invariance is not clear: the data spans too small a range to rule out other heavy-tailed distributions like the log normal. Based on the nearly three decades of relatively clean scaling for the United States, however, we conjecture that the formation of road networks, when conducted at a sufficiently large scale, leads to true scale-free structure.

We fit these distributions using a maximum likelihood estimator for the power law, as in [20]. For the United States, England, and Denmark, we find scaling exponents of $\alpha = 2.4$, 2.2, and 2.4, respectively. These are likely slight overestimates of their true values; progressive subsampling of each network yields slightly larger estimates of α. Using Monte Carlo simulation and the Kolmogorov-Smirnov goodness-of-fit test, in the manner of [21], we find that the models are good representations of the data, with $p_{MC} \geq 0.99$. Finally, we do not propose that α has a universal value. Rather, in the next section, we describe a toy model which can give rise to a variety of exponents, depending on the fractal dimensions describing the placement of roads and intersections.

VI. A SIMPLE FRACTAL MODEL

In this section we introduce and analyze a simple fractal model for the placement of roads on the unit square that
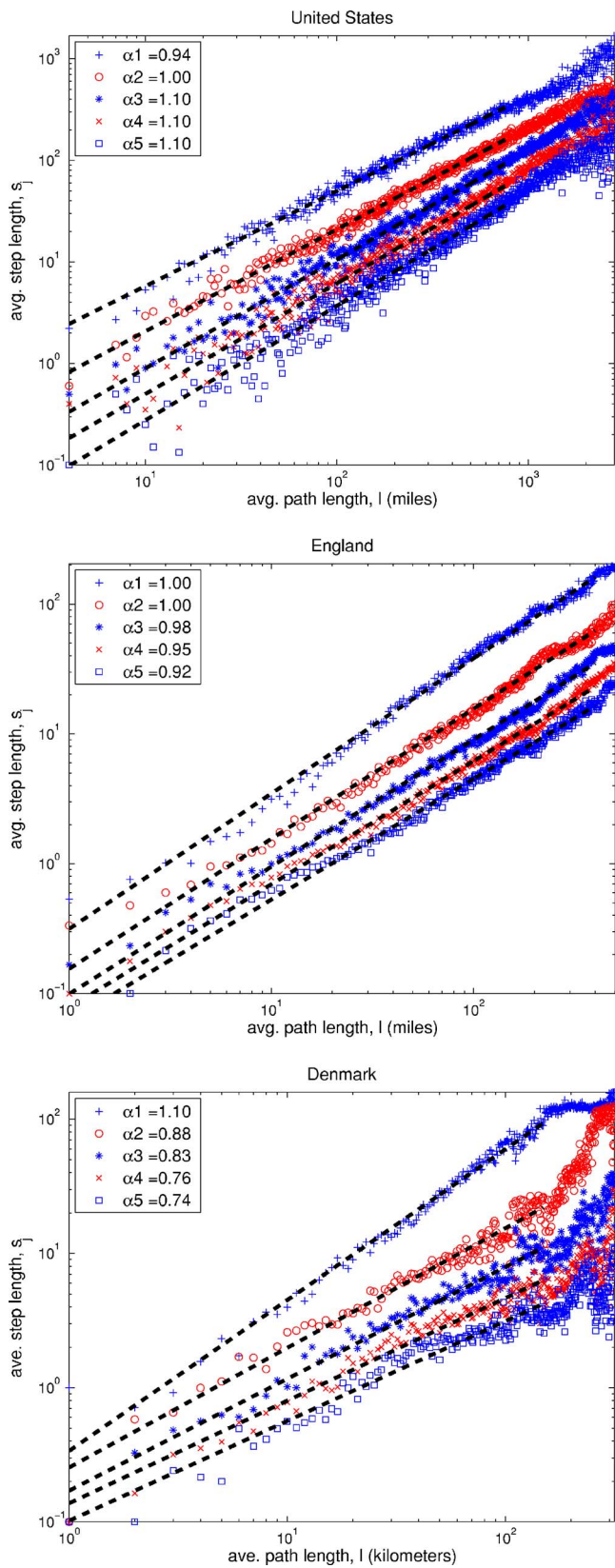
FIG. 2. (Color online) The scale invariant hypothesis predicts that $s_j \approx A_j \ell$ for constants A_j, and thus that $\alpha_j \approx 1$. This is consistent with our power-law fits, in which we estimate α_j using a bootstrap resampling method. Journeys on the very largest scales were excluded in order to avoid finite-size effects.
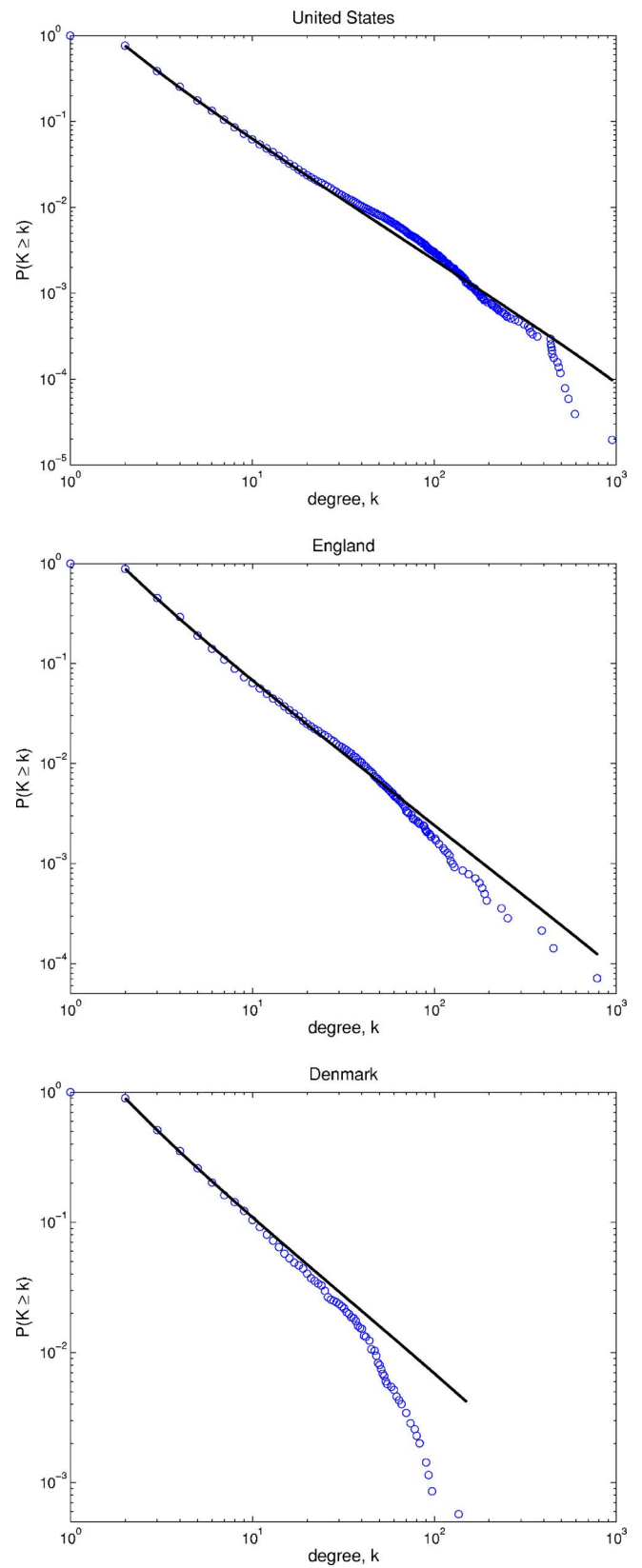
FIG. 3. (Color online) The cumulative degree distributions $P(K \geq k)$ of the dual model for the United States, England, and Denmark. We show fits based on a maximum likelihood estimate of a power law model $P(k) \sim k^{-\alpha}$, where $\alpha = 2.4$, 2.2, and 2.4, respectively.
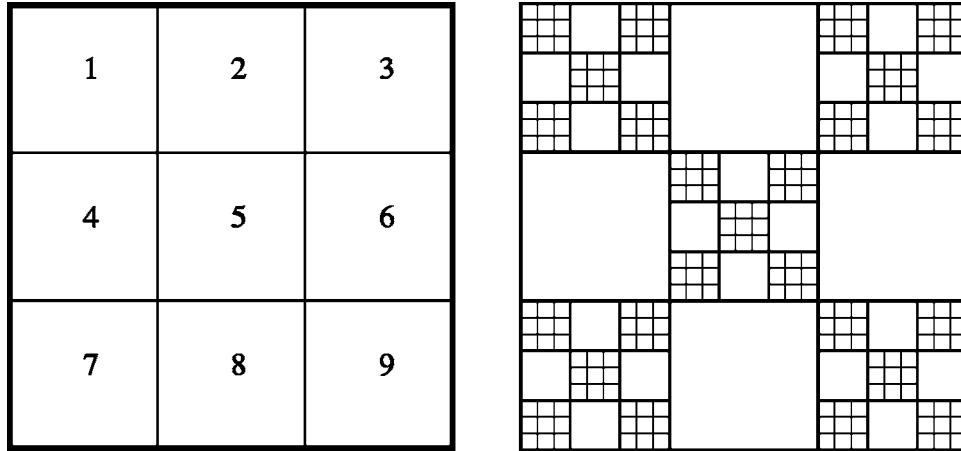
FIG. 4. A version of our fractal model for road placement (see text). Line-thickness indicates greater road capacity and speed limits. The Sierpinski carpet corresponds to recursively subdividing all squares except square 5.

reproduces both the observed hierarchical and scale invariant structure of journeys. As we will see, the key quantities of the model are the fractal or Hausdorff dimensions d_p and d_i that, in turn, describe the distribution of road intersections in the plane, and the distribution of intersections along a single road.

Unlike previous models of physical networks [11–13], our model assumes no optimization or resource constraint satisfaction mechanism. Rather, we simply assume the fractal structure is given, and analyze the resulting implications for journey structure and the dual degree distribution. We leave for future work the exploration of mechanisms that may in turn generate a fractal placement of roads.

To create a road network according to our model, we first divide the unit square into η^2 squares of equal size for some fixed integer η by placing $2(\eta-1)$ roads. We then choose some subset of these η^2 squares and subdivide them as we did the original square, by placing $2(\eta-1)$ new roads per subdivision. Repeating this process recursively for as many levels as desired yields a road network with fractal structure, where lines are roads and line crossings are intersections. For instance, with $\eta=3$, subdividing all but the center square gives the Sierpinski carpet [22], and in Fig. 4 we show a network resulting from subdividing five of the nine squares.

Observe that in this model the road intersections are distributed as a fractal both over the original unit square and along a given road. For instance, in the Sierpinski carpet, at each level of construction the total number of intersections in the plane increases by a factor of 8, while the number of intersections along a given road triples; this construction thus yields a fractal dimension $d_p = \log_3 8$ for the distribution in the plane, and $d_i = \log_3 3 = 1$ for the distribution along a given road. Similarly, for the scheme illustrated in Fig. 4, at each level the number of intersections increases by a factor of 5 and the number of intersections along a given road doubles, giving $d_p = \log_3 5$ and $d_i = \log_3 2$.

We show, by a simple counting argument, that the scaling exponent of such a network's dual degree distribution is related to the fractal dimensions in the following way:

$$\alpha = 1 + \frac{d_p}{d_i}. \qquad (2)$$

For each $x \leq m$, where m is the total number of levels of subdivision, the number of roads at level x is $r(x) \sim \eta^{d_p x}$. Similarly, for a road added at level x, the number of intersections along its length is exponential in the number of subsequent subdivisions, and is given by $c(x) \sim \eta^{d_i(m-x)} \sim \eta^{-d_i x}$.

The cumulative degree distribution of this model can be calculated as follows. The number of roads with degree greater than k is given by

$$P(K > k) = \sum_{x:c(x)>k} r(x) \sim \sum_{x=m-(1/d_i)\log_\eta k}^{m} \eta^{d_p x} \sim k^{-d_p/d_i}.$$

So, differentiating this cumulative distribution gives the degree distribution $P(k) \sim k^{-\alpha}$, with α given by Eq. (2). The values of α for a few variations on the $\eta=3$ subdivision schema are given in Table II.

Further, by placing roads hierarchically through the subdivision process, journeys that seek to minimize travel time will necessarily utilize this same hierarchical structure, especially if roads at the earlier levels of construction correspond to roads with higher speed limits and traffic capacities. For instance, if the source and destination are in different subsquares, then the shortest path in the dual model will use one

TABLE II. Fractal dimensions for the distribution of intersections in the plane d_p, the distribution of intersections along a single road d_i, and the power-law exponent α for different subdivision schemes given by Eq. (2) for $\eta=3$ (see text).

Schema	d_p	d_i	α
all	$\log_3 9$	$\log_3 3$	3.00
all but center	$\log_3 8$	$\log_3 3$	2.89
odd numbers	$\log_3 5$	$\log_3 2$	3.32
corners	$\log_3 4$	$\log_3 2$	3.00

of the roads at level $x=1$; this is also recursively true at each step of the journey. Thus, the jth largest step will cover an average fraction A_j of the journey, which scales as $A_j \sim n^{-j}$. Indeed, looking at the data for the United States (Table I), it appears that A_j decreases roughly exponentially with j.

The fact that our toy fractal model reproduces the scale invariant journey structure, and can similarly produce the correct functional form of the dual degree distribution, suggests that the roads in our real world networks may be organized in a similar fractal structure. It would be interesting to use the geographic distribution of population and road intersections to estimate the fractal dimensions d_p and d_i for various countries, and compare the value of α predicted by Eq. (2) to the measured value. We leave this as a direction for future work.

VII. CONCLUSION

We studied the national road networks of the United States, England, and Denmark through their dual representation, using the driving directions provided by a popular commercial service. Like those of urban road networks [16], we found that the dual degree distribution is characterized by a heavy tail; however, for large countries such as the United States, this distribution is likely scale free, following a power law of the form $P(k) \sim k^{-\alpha}$, with $2.2 \le \alpha \le 2.4$ (Fig. 3).

We further showed that journeys on these networks have a scale invariant structure, characterized by a driver rising up through the road hierarchy, i.e., from local to regional to national roads where the speed limit and capacity grows with each step, and then descending in reverse order as she approaches her destination. This scale invariance is exhibited by the fact that journeys have similar structure regardless of their total length. Notably, our empirical observations say nothing about the actual traffic density on these roads, which is likely determined by the nonuniform popularity of destinations.

To explain the observed structure in the road networks, we introduced and analyzed a simple fractal model of road placement. This model recovers the scale-free structure of journeys in the network and the power-law dual degree distribution. It also suggests a fundamental relationship between the exponent α and the fractal dimensions describing the distribution of road intersections in the plane d_p and along a single road d_i. Although our model assumes that road placement is not a function of resource-bound optimization as in [12,13], it would be interesting to adapt it in such a way as to generate more statistically realistic road networks. Arguably, biological transportation networks, e.g., vascular networks, also have a fractal structure [23], and a comparative study of these and our road networks would be interesting.

ACKNOWLEDGMENTS

[1] R. Albert and A.-L. Barabasi, Rev. Mod. Phys. **74**, 47 (2002).

[2] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).

[3] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).

[4] M. Faloutsos, P. Faloutsos, and C. Faloutsos, Comput. Commun. Rev. **29**, 251 (1999).

[5] J. M. Kleinberg, S. R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, in *Proceedings of the International Conference on Combinatorics and Computing*, Lecture Notes in Computer Science No. 1627 (Springer-Verlag, Berlin, 1999), pp. 1–18.

[6] S. Wasserman and K. Faust, *Social Network Analysis* (Cambridge University Press, Cambridge, 1994).

[7] D. J. de S. Price, Science **149**, 510 (1965).

[8] S. Redner, e-print physics/0407137.

[9] T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, and Y. Sakaki, Proc. Natl. Acad. Sci. U.S.A. **98**, 4569 (2001).

[10] J. M. Montoya and R. V. Solé, J. Theor. Biol. **214**, 405 (2002).

[11] M. T. Gastner and M. E. J. Newman, e-print cond-mat/0407680.

[12] A. Fabrikant, E. Koutsoupias, and C. H. Papadimitriou, in *Proceedings of International Colloquia on Automata, Languages and Programming*, Lecture Notes in Computer Science No. 2380 (Springer-Verlag, Berlin, 2002), pp. 110–122.

[13] S. S. Manna and A. Kabakçioglu, J. Phys. A **36**, 19 (2003).

[14] S. Porta, P. Crucitti, and V. Latora, e-print cond-mat/0411241.

[15] B. Jiang and C. Claramunt, Environ. Plan. B: Plan. Des. **31**, 151 (2004).

[16] M. Rosvall, A. Trusina, P. Minnhagen, and K. Sneppen, Phys. Rev. Lett. **94**, 028701 (2005).

[17] S. Porta, P. Crucitti, and V. Latora, e-print cond-mat/0506009.

[18] We note that while not everyone navigates in this way, this is the form of the results returned by the increasingly popular navigational services from companies such as Google, Yahoo!, Microsoft, and Mapquest.

[19] W. Hillier, www.spacesyntax.org/publications/commonlang.html.

[20] A. Clauset and M. Young, e-print physics/0502014.

[21] M. L. Goldstein, S. A. Morris, and G. G. Yen, Eur. Phys. J. B **41**, 255–258 (2004).

[22] M. Schroeder, *Fractals, Chaos, Power Laws: Minutes From an Infinite Paradise* (Freeman, New York, 1992).

[23] G. B. West, J. H. Brown, and B. J. Enquist, Science **284**, 1667 (1999).