Scale Invariance in Road Networks

Vamsi Kalapala[†], Vishal Sanwalani[†], Aaron Clauset[†,*] and Cristopher Moore[†,‡]
[†]*Department of Computer Science and* [‡]*Department of Physics and Astronomy, University of New Mexico, Albuquerque, NM 87131.*
`(vamsi, vishal, aaron, moore)@cs.unm.edu`

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 empirically show that the degree distribution of the dual is well-characterized by a power law with exponent $2.0 < \alpha < 2.5$, 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 connection between the scaling exponent α and the fractal dimensions governing the placement of roads and intersections.

I. INTRODUCTION

The study of 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 citation [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. For example, the length of a path between two vertices may be either the number of edges to cross or the sum of the lengths of those edges; notably, the shortest path in one sense need not be the shortest path in the other sense. 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 contiguous road of a given name, and two vertices are joined if their corresponding roads ever intersect. This graph transformation has been used previously to study the topological structure of urban roads [14–17]. This should not to be confused with the dual of a planar graph, in which faces become vertices and vice versa.

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 $2.0 < \alpha < 2.5$. 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.

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 $k = 4$, and the average degree of a planar network is at most $\langle k \rangle < 6$. 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 [22], 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 primal road network, we must define which road segments naturally belong together. In previous studies of road networks, segments have been

*Corresponding author.

grouped by their street name [15, 16], line of sight by a driver [18], 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 sample 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 a pair of source and destination addresses. If only partial information is provided, e.g., the postal code or city name, the service defaults to a unique address near that location's center. For a pair of addresses, the service then returns a driving directions as a list of road names, the respective distances a driver should travel on each one, and instructions as to how to get from one road to another, e.g., "turn left onto" or "continue on".

We constructed samples of each nation's respective road network by taking the union of the paths returned for a large number source-destination pairs. For the United States, we selected 200 000 uniformly random pairs of postal codes, while for England and Denmark, we used 25 000 uniformly random pairs of city names.

Notably, our sampled networks are biased according to population distribution (postal codes in the United States are distributed roughly according to population [23]). On the other hand, by focusing only on travel between postal codes for the United States, and between cities in England and Denmark, 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 creates 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 will classify 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 possiblew, and then descends to progressively smaller roads until she reaches the local street of her destination.

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 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 regional highway off-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.

In Table I, we show data for the five largest steps of these journeys; these steps alone typically account for 85% or more of the total length of the journey, and in the United States, the largest step typically covers about 40% of the entire distance, the second largest covers 20%, the third largest covers 12%, etc. Moreover, for each j from 1 to 5, the fraction of the journey covered by the jth largest step appears to be constant. This suggests a simple linear relationship of the form

$$s_j = A_j \, \ell \ , \tag{1}$$

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 quite well; moreover, we have $\alpha_j \approx 1$, suggesting that the linear form of (1) is correct.

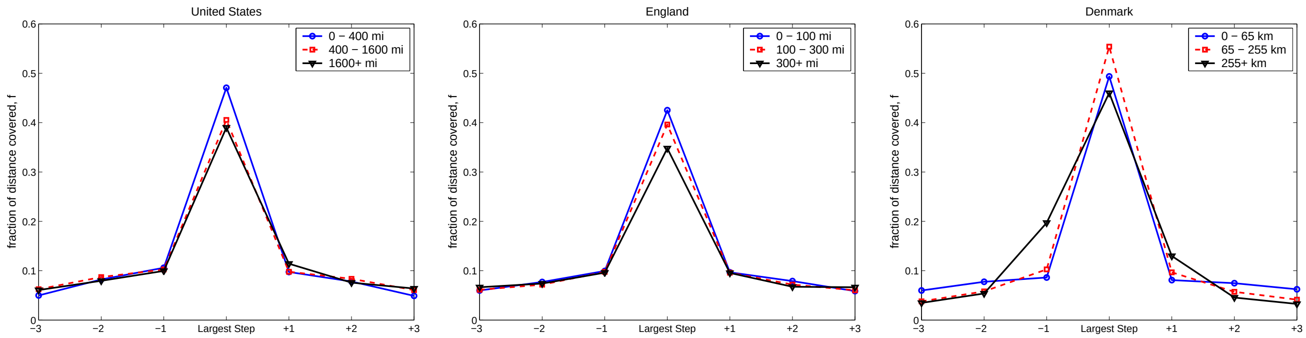

FIG. 1: (Color online) The average journey profiles for the United States, England and Denmark. Profiles are composed of the largest step (centered) and the three (in order of appearance) largest steps which precede and follow it.

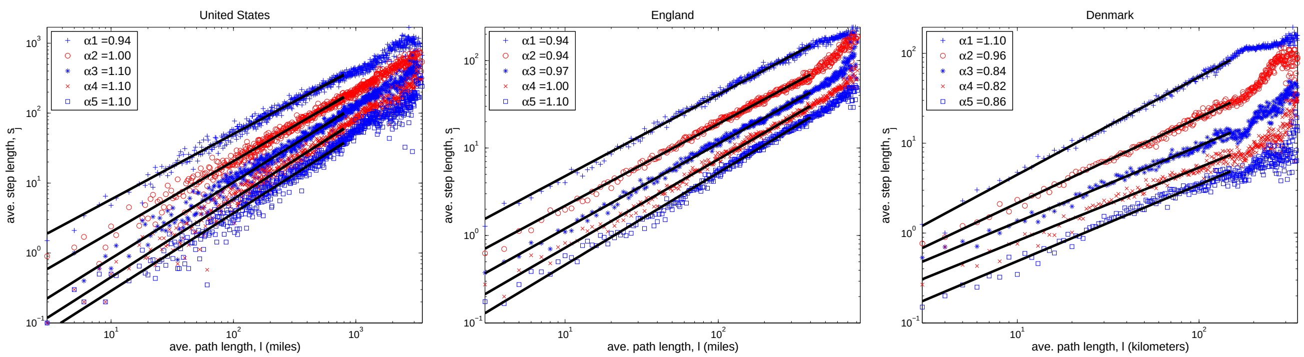

FIG. 2: (Color online) The scale invariant hypothesis predicts that $s_j \approx A_j\,\ell$ for constants A_j, and thus that $\alpha_j \approx 1$. This is consistent with our power-law fits, in which we estimate α_j using a bootstrap resampling method. Journeys on the very largest scales were excluded in order to avoid finite-size effects.

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, shown in Figure 3, although with apparent finite-size cutoffs related to the respective geographic scales. A power law model of these distributions is most convincing for the United States, where the power law is quite clean for almost three decades. For England and Denmark the situation is less clear (England in particular shows strong curvature on the log-log plot). We conjecture that this is because the English and Danish networks have fewer levels of hierarchy than their American counterpart, and that the formation of road networks, when conducted at a sufficiently large scale, leads to scale-free structure.

We fit these distributions using the maximum likelihood estimator for the one-parameter power law, as in [19]. For the United States, England and Denmark, we find scaling exponents of $\alpha = 2.4, 2.1$ and 2.4, respectively. We do not propose that α has a universal value; in fact, 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 roads and their 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 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 two dimensions, 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

Network	Distance		Largest Step	2^{nd} Largest Step	3^{rd} Largest Step	4^{th} Largest Step	5^{th} Largest Step	Sum
	0 - 400		0.486 ± 0.184	0.212 ± 0.087	0.112 ± 0.057	0.066 ± 0.040	0.042 ± 0.030	0.917
United States	400 - 1600	mi	0.411 ± 0.161	0.209 ± 0.074	0.126 ± 0.051	0.081 ± 0.039	0.053 ± 0.030	0.880
	1600+		0.391 ± 0.139	0.212 ± 0.067	0.128 ± 0.048	0.079 ± 0.035	0.053 ± 0.027	0.862
	0 - 100		0.432 ± 0.168	0.201 ± 0.075	0.117 ± 0.051	0.074 ± 0.037	0.048 ± 0.028	0.873
England	100 - 300	mi	0.397 ± 0.172	0.184 ± 0.066	0.111 ± 0.042	0.076 ± 0.032	0.055 ± 0.027	0.823
	300+		0.347 ± 0.106	0.179 ± 0.065	0.110 ± 0.028	0.076 ± 0.017	0.056 ± 0.017	0.768
	0 - 65		0.499 ± 0.183	0.199 ± 0.083	0.110 ± 0.056	0.067 ± 0.039	0.043 ± 0.028	0.918
Denmark	65 - 250	km	0.562 ± 0.179	0.199 ± 0.095	0.089 ± 0.054	0.049 ± 0.033	0.031 ± 0.025	0.931
	250+		0.463 ± 0.069	0.295 ± 0.114	0.103 ± 0.042	0.050 ± 0.010	0.027 ± 0.020	0.938

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.

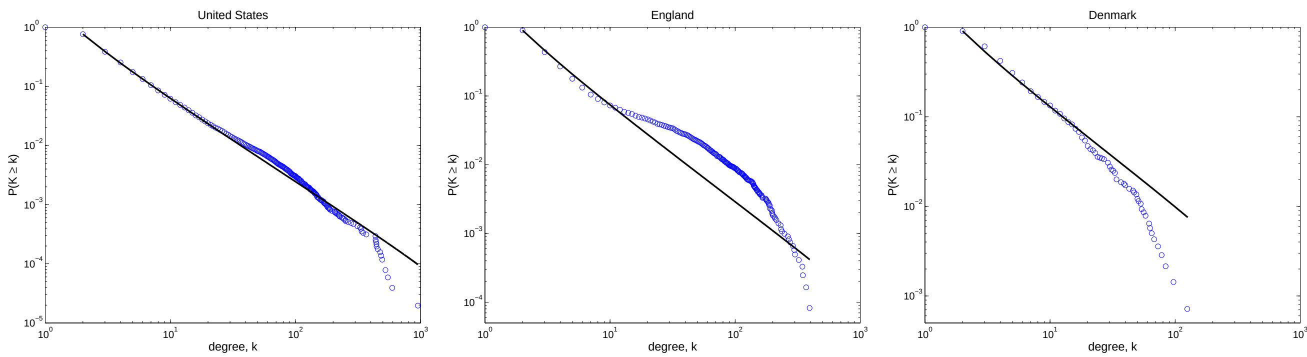

FIG. 3: (Color online) The cumulative degree distributions $P(K \geq k)$ of the dual model for the United States, England, and Denmark. We show fits based on a maximum likelihood estimate of a power law model $P(k) \sim k^{-\alpha}$, where $\alpha = 2.4, 2.1$ and 2.4, respectively.

first divide the unit square into n^2 squares of equal size for some fixed integer n by placing $2(n-1)$ roads. We then choose some subset of these n^2 squares and subdivide them as we did the original square. Repeating this process recursively for as many levels as desired yields a road network with fractal structure. For instance, with $n = 3$, subdividing all but the center square gives the Sierpinski carpet [20], and in Figure 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 Figure 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,

$$\alpha = 1 + \frac{d_p}{d_i} \quad . \tag{2}$$

After x levels of subdivision, the number of roads at level x is proportional to the number of squares at that level, and so grows exponentially with x as $r(x) \sim n^{d_p x}$. Similarly, for a road added at level x, the number of intersections along its length is exponential in the number of subsequent subdivisions, and so its degree grows as $c(x) \sim n^{d_i(m-x)} \sim n^{-d_i x}$, where m is the total number of levels of subdivision.

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

$$\begin{aligned} P(K > k) &= \sum_{x \,:\, c(x) > k} r(x) \\ &\sim \sum_{x = m - (1/d_i)\log_n k}^{m} n^{d_p x} \\ &\sim k^{-d_p/d_i} \quad . \end{aligned}$$

So, differentiating this cumulative distribution gives the degree distribution $P(k) \sim k^{-\alpha}$, with α given by Eq. (2).

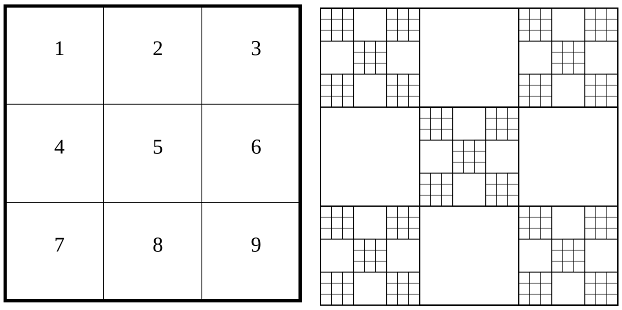

FIG. 4: A version of our fractal model for road placement (see text). Line-thickness indicates greater road capacity and speed limits. The Sierpinski carpet corresponds to recursively subdividing all squares except square 5.

The values of α for variations on $n = 3$ 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. For instance, if the source and destination are in different subsquares, then the shortest path in the dual model will use one 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, while scales as $A_j \sim n^{-j}$. Indeed, looking at the data for the United States (Figure 1), it appears that A_j decreases roughly exponentially with j.

The fact that our simple 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.

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
odds	$\log_3 5$	$\log_3 2$	3.32
corners	$\log_3 4$	$\log_3 2$	3.00

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 $n = 3$ (see text).

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. We show that the dual degree distributions have heavy tails, like those of urban road networks [16], and are well modeled by power laws of the form $P(k) \sim k^{-\alpha}$, with $2.0 < \alpha < 2.5$ (Fig. 3), like many other real world networks [1–3].

We further show that journeys on these networks follows a scale invariant structure, in which a driver rises 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 descends 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.

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 and along a single road. 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., the circulatory system, also have a fractal structure [21], and a comparative study of these and our road networks would be interesting.

Acknowledgments

The authors thank Sahar Abubucker for reviewing the manuscript and Vincent Nesme for conversations on Hausdorff dimensions.This work was supported by NSF grants PHY-0071139, PHY-0200909, and the Sandia University Research Project.

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

[2] S. N. Dorogovtsev and J. F. F. Mendes, Evolution of networks. *Advances in Physics* **51**, 1079 (2002).

[3] M. E. J. Newman, The structure and function of complex networks. *SIAM Review* **45**, 167 (2003).

[4] M. Faloutsos, P. Faloutsos, and C. Faloutsos, On power-law relationships of the internet topology. *Computer Communications Review* **29**, 251 (1999).

[5] J. M. Kleinberg, S. R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, "The Web as a graph: Measurements, models and methods," in *Proceedings of the International Conference on Combinatorics and Computing*, No. 1627 in Lecture Notes in Computer Science (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, Networks of scientific papers. *Science* **149**, 510 (1965).

[8] S. Redner, Citation Statistics From More Than a Century of Physical Review. e-print `physics/0407137` (2004)

[9] T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, and Y. Sakaki, A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4569 (2001).

[10] J. M. Montoya and R. V. Solé, Small world patterns in food webs. *J. Theor. Bio.* **214**, 405 (2002).

[11] M. T. Gastner, M. E. J. Newman, Spatial Structure of Networks, e-print `cond-mat/0407680` (2004).

[12] A. Fabrikant, E. Koutsoupias and C. H. Papadimitriou, "Heuristically optimized trade-offs: a new paradigm for power-laws in the internet," in *Proceedings of International Colloquia on Automata, Languages and Programming*, No. 2380 in Lecture Notes in Computer Science (Springer-Verlag, Berlin, 2002), pp. 110-122.

[13] S. S. Manna and A. Kabakçioglu, Scale-free network on Euclidean space optimized by Rewiring of Links. *J. Phys. A: Math. Gen.* **36**, 19 (2003).

[14] S. Porta, P. Crucitti and V. Latora, The Network Analysis of Urban Streets : A dual approach. e-print, `cond-mat/0411241` (2004).

[15] B. Jiang and C. Claramunt, Topological Analysis of Urban Street Networks, *Environment and Planning B* **31**, 151 (2004).

[16] M. Rosvall, A. Trusina, P. Minnhagen and K. Sneppen, Networks and Cities: An Information Perspective. *Phys. Rev. Lett.* **94**, 028701 (2005).

[17] S. Porta, P. Crucitti and V. Latora, The Network Analysis of Urban Streets : A primal approach. e-print, `cond-mat/0506009` (2005)

[18] W. Hillier, The Common Language of Space, `www.spacesyntax.org/publications/commonlang.html`.

[19] A. Clauset and M. Young, Scale Invariance in Global Terrorism. e-print, `physics/0502014` (2005).

[20] M. Schroeder, *Fractals, Chaos, Power Laws : Minutes from an infinite paradise.* (W.H. Freeman, New York, 1992).

[21] G. B. West, J. H. Brown and B. J. Enquist, The Fourth Dimension of Life: Fractal Geometry and Allometric Scaling of Organisms. *Science* **284**, 1667 (1999).

[22] 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 and Mapquest.

[23] See, for instance, `www.census.gov/geo/www/gazetteer/gazette.html`