Doctopic: Statistical physics ARTICLE IN PRESS PLA:24496

Physics Letters A ••• (••••) •••–•••

Contents lists available at ScienceDirect

Physics Letters A

www.elsevier.com/locate/pla

Non-monotonicity of the generalized dimensions of a complex network

Eric Rosenberg

AT&T Labs, Middletown, NJ 07748, USA

ARTICLE INFO

Article history:
Received 26 March 2017
Received in revised form 7 May 2017
Accepted 8 May 2017
Available online xxxx
Communicated by C.R. Doering

Keywords:
Complex networks
Generalized dimensions
Multifractals
Box counting

ABSTRACT

Computing the generalized dimensions D_q of a complex network requires covering the network by a minimal number of "boxes" of size s, for a range of s. We show that, unlike the case for a geometric multifractal, for a complex network the shape of the D_q vs. q curve can be monotone increasing, or monotone decreasing, or even have both a local maximum and a local minimum, depending on the range of box sizes used to compute D_q. We provide insight into this behavior by deriving a simple closed-form expression for the derivative of D_q at $q = 0$. The estimate depends on the ratio of the geometric mean of the box masses (where the mass of a box is the number of nodes it contains) to the arithmetic mean of the box masses.

1. Introduction

A multifractal is a fractal that cannot be characterized by a single fractal dimension such as the box counting dimension. The infinite number of fractal dimensions needed to characterize a multifractal are known as generalized dimensions. Generalized dimensions of geometric multifractals were proposed independently in 1983 by Grassberger [7] and by Hentschel and Procaccia [9]. They have been intensely studied, e.g., [17] and widely applied (e.g., [16, 26]). Given N points from a geometric multifractal (e.g., the strange attractor of a dynamical system [18]), the generalized dimension D_q is computed by covering the N points with a grid of boxes of linear size s, computing the fraction $p_j(s)$ of the N points in box B_j of the grid, computing the partition function $Z_q(s) = \sum_j [p_j(s)]^q$ (where the sum is over all boxes in the grid), and examining how $\log Z_q(s)$ scales with $\log s$. When $q = 0$ this computation yields the box counting dimension d_B, so $D_0 = d_B$. When $q = 1$ we obtain the information dimension d_I [3], so $D_1 = d_I$. When $q = 2$ we obtain the correlation dimension d_C [8], so $D_2 = d_C$.

Complex networks have also been studied from a multifractal perspective. A network $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ is a set $\mathcal{N}$ of nodes connected by a set $\mathcal{A}$ of arcs. For example, in a friendship social network, a node might represent a person and an arc indicates that two people are friends. Many applications of network models are discussed in [2]. We use the term "complex network" to mean an arbitrary network without special structure (as opposed to, e.g., a regular lattice), for which all arcs have unit cost (so the length of a shortest path between two nodes is the number of arcs in that path), and all arcs are undirected (so the arc between nodes i and j can be traversed in either direction). The box counting dimension of a complex network was proposed in ([12,24,25]), the information dimension was studied in ([22,29]), the correlation dimension was studied in ([13,21]), and generalized dimensions were studied in ([4,23,27,28]).

In [22], Rosenberg showed that the definition proposed in [29] of the information dimension d_I of a complex network $\mathcal{G}$ is ambiguous, since d_I is computed from a minimal covering of $\mathcal{G}$ by "boxes", and there are in general different minimal coverings of $\mathcal{G}$ by boxes of size s, and these different minimal coverings yield different values of d_I. The ambiguity is resolved for each s by maximizing the entropy over the set of minimal coverings by boxes of size s. The same ambiguity is present when using box counting to compute D_q (as in [27]), since the different minimal coverings by boxes of size s can yield different values of D_q. The solution proposed in [23] is to select, for each s, the minimal covering yielding the lexico (short for "lexicographically") minimal summary vector $x(s)$. The value $x_j(s)$ is the number of nodes in box B_j of the minimal covering, and $x_j(s)$ is non-increasing in j. For each s, there is a unique $x(s)$, and for all sufficiently large q the minimal covering summarized by $x(s)$ minimizes the partition function $Z_q(s)$ over the set of minimal coverings of size s. Computing $x(s)$ requires negligible extra computation beyond what is required to compute a minimal covering. The $x(s)$ vectors can be used to unambiguously compute D_q for $q \neq 1$.

E-mail address: ericr@att.com.

http://dx.doi.org/10.1016/j.physleta.2017.05.014

Whereas for geometric multifractals it is known [7] that D_q is monotone non-increasing in q, here we show that for a complex network, even when D_q is computed using the lexico minimal summary vectors $x(s)$, the D_q vs. q curve is *not* necessarily monotone non-increasing, and this curve can assume different shapes, depending on the range of box sizes used to compute D_q. To see how the D_q vs. q curve can assume different shapes for the same complex network $\mathcal{G}$, consider how D_q is computed. The computation is based upon the scaling of $Z_q(s) = \sum_j [p_j(s)]^q$, where $p_j(s)$ is the fraction of nodes in box B_j of a minimal covering of $\mathcal{G}$ by boxes of size s. For a given q, we identify a range $[L, U]$ of box sizes, where $L < U$, such that $\log Z_q(s)$ is approximately linear in $\log s$ for $s \in [L, U]$. The estimate of D_q is $1/(q-1)$ times the slope of the linear approximation [23]. While in theory L and U could depend on q, in practice, as in this paper, we assume that a single choice of L and U is used to compute D_q for all q. Rather than estimating D_q using a technique such as regression over the range $[L, U]$ of box sizes, we instead estimate D_q using only the two box sizes L and U. Such a two-point estimate was used in [21], where it was shown that even for as simple a network as a one-dimensional chain of nodes, the two-point estimate has very desirable properties. From an analytical point of view, the huge benefit of the two-point estimate is that it yields a closed-form estimate of D_q; we denote this estimate by $D_q(L, U)$.

We plot $D_q(L, U)$ vs. q for several networks, and show that the $D_q(L, U)$ vs. q curve can assume dramatically different shapes, depending on the choice of L and U: the curve can be monotone increasing, or monotone decreasing, or have a local minimum or a local maximum, or even have both a local minimum and a local maximum. (Such behavior stands in sharp contrast to the behavior of a geometric multifractal, for which the curve is monotone non-increasing.) We provide insight into this behavior by deriving a simple closed-form expression for $D_0'(L, U)$, the derivative of $D_q(L, U)$ at $q = 0$. Interestingly, this derivative depends on the ratio of the geometric mean of the box masses to the arithmetic mean of the box masses. We discuss the relationship of this ratio to the maximal entropy criterion [22] and the lexico minimal criterion [23] used to compute d_I and D_q, respectively. The theoretical results we present suggest there is a rich theory of the generalized dimensions of a complex network, and in Section 6 we suggest some areas for investigation. For convenience, the symbols used in this paper are summarized in Table 1.

Table 1
Symbols and their definitions.

Symbol	Definition
Δ	network diameter
$A(s)$	arithmetic mean of the box masses in $\mathcal{B}(s)$
$\mathcal{B}(s)$	covering of $\mathcal{G}$ of size s
$B(s)$	cardinality of $\mathcal{B}(s)$
B_j	box in $\mathcal{B}(s)$
d_B	box counting dimension
D_q	generalized dimension of order q
$D_q(L, U)$	secant estimate of D_q
d_I	information dimension
$\mathcal{G}$	complex network
$G(s)$	geometric mean of the box masses in $\mathcal{B}(s)$
$H(s)$	entropy of the distribution $p_j(s)$
N	number of nodes in $\mathcal{G}$
$N_j(s)$	number of nodes in box $B_j \in \mathcal{B}(s)$
$p_j(s)$	probability of box $B_j \in \mathcal{B}(s)$
$\mathbb{R}^E$	E-dimensional Euclidean space
$x(s)$	lexico minimal vector summarizing $\mathcal{B}(s)$
$Z_q(\mathcal{B}(s))$	partition function value for the covering $\mathcal{B}(s)$
$Z(x, q)$	partition function value for the summary vector x

2. Preliminary definitions

Throughout this paper, $\mathcal{G}$ will refer to a complex network with node set $\mathcal{N}$ and arc set $\mathcal{A}$. We assume that $\mathcal{G}$ is connected, meaning there is a path of arcs in $\mathcal{A}$ connecting any two nodes. Let $N \equiv |\mathcal{N}|$ be the number of nodes in $\mathcal{G}$. The network B is a subnetwork of $\mathcal{G}$ if B is connected and B can be obtained from $\mathcal{G}$ by deleting nodes and arcs. For each positive integer s such that $s \geq 2$, let $\mathcal{B}(s)$ be a collection of subnetworks (called *boxes*) of $\mathcal{G}$ satisfying two conditions: (i) each node in $\mathcal{N}$ belongs to exactly one subnetwork (i.e., to one box) in $\mathcal{B}(s)$, and (ii) the diameter of each box in $\mathcal{B}(s)$ is at most $s-1$. (The diameter Δ of a network is defined by $\Delta \equiv \max\{dist(x, y) \,|\, x, y \in \mathcal{N}\}$, where $dist(x, y)$ is the length of the shortest path between nodes x and y, and where "$\equiv$" denotes a definition.) We call $\mathcal{B}(s)$ a *covering* of $\mathcal{G}$ of size s, or more simply, an s-covering. We do not consider $\mathcal{B}(s)$ for $s = 1$, since a box of diameter 0 contains only a single node. Define $B(s) = |\mathcal{B}(s)|$, so $B(s)$ is the number of boxes in $\mathcal{B}(s)$. The s-covering $\mathcal{B}(s)$ is *minimal* if $B(s)$ is less than or equal to the number of boxes in any other s-covering. For $s > \Delta$, the minimal s-covering $\mathcal{B}(s)$ consists of a single box which is $\mathcal{G}$ itself. The term "box counting" refers to computing a minimal s-covering for a range of values of s. In general, we cannot easily compute a minimal s-covering, but good heuristics are known (e.g., [5,12,19,24]).

Let $\mathcal{B}(s)$ be an s-covering of $\mathcal{G}$. For $B_j \in \mathcal{B}(s)$, define $p_j(s) \equiv N_j(s)/N$, where $N_j(s)$ is the number of nodes in B_j. By the "mass" of a box B_j, we mean the number of nodes in B_j. For $q \in \mathbb{R}$, define

$$Z_q(\mathcal{B}(s)) \equiv \sum_{B_j \in \mathcal{B}(s)} [p_j(s)]^q \,. \tag{1}$$

We call $Z_q(\mathcal{B}(s))$ the *partition function value* for $\mathcal{B}(s)$. Following [23], we summarize $\mathcal{B}(s)$ by the point $x(s) \in \mathbb{R}^J$, where $J = B(s)$, where $x_j(s) = N_j(s)$ for $1 \leq j \leq J$, and where $x_1(s) \geq x_2(s) \geq \cdots \geq x_J(s)$. We say "summarize" since $x(s)$ does not specify all the information in $\mathcal{B}(s)$; in particular, $\mathcal{B}(s)$ specifies exactly which nodes belong to each box, while $x(s)$ specifies only the mass of each box. We use the notation $x(s) = \sum \mathcal{B}(s)$ to mean that $x(s)$ summarizes the s-covering $\mathcal{B}(s)$ and that $x_1(s) \geq x_2(s) \geq \cdots \geq x_J(s)$. For example, if $N = 37$, $s = 3$, and $B(3) = 5$ then we might have $x(3) = \sum \mathcal{B}(3)$ for $x(3) = (18, 7, 5, 5, 2)$. If $x(s) = \sum \mathcal{B}(s)$ then each $x_j(s)$ is positive, since $x_j(s)$ is the mass of box B_j.

Let $x \in \mathbb{R}^K$ for some positive integer K. Let $right(x) \in \mathbb{R}^{K-1}$ be the point obtained by deleting the first component of x. For example, if $x = (18, 7, 5, 5, 2)$ then $right(x) = (7, 5, 5, 2)$. Let $u \in \mathbb{R}$ and $v \in \mathbb{R}$ be numbers. We say that $u \succeq v$ (in words, u is *lexico* greater than or equal to v) if ordinary inequality holds, that is, $u \succeq v$ if $u \geq v$. (We use *lexico* instead of the longer *lexicographically*.) Now let $x \in \mathbb{R}^K$ and $y \in \mathbb{R}^K$. We define lexico inequality recursively. We say that $y \succeq x$ if either (i) $y_1 > x_1$ or (ii) $y_1 = x_1$ and $right(y) \succeq right(x)$. Thus $(9, 6, 5) \succeq (8, 7, 5)$ and $(9, 6, 7) \succeq (9, 5, 8)$.

Definition 1. Let $x(s) = \sum \mathcal{B}(s)$. Then $x(s)$ is *lexico minimal* if (i) $\mathcal{B}(s)$ is a minimal s-covering and (ii) if $\widetilde{\mathcal{B}}(s)$ is a minimal s-covering distinct from $\mathcal{B}(s)$ and $\widetilde{x}(s) = \sum \widetilde{\mathcal{B}}(s)$ then $\widetilde{x}(s) \succeq x(s)$. ■

It is proved in [23] that for each s there is a unique lexico minimal summary. Moreover, if $x(s) = \sum \mathcal{B}(s)$ is lexico minimal then for all sufficiently large q and for any other minimal s-covering $\widetilde{\mathcal{B}}(s)$ we have $Z_q(\mathcal{B}(s)) \leq Z_q(\widetilde{\mathcal{B}}(s))$. Since for $q > 1$ the partition