Chapter 10
Non-monotonicity of Generalized Dimensions

In Chap. 9, we showed that the value of D_q for a given q depends in general on which minimal s-covering is selected, and we showed that this ambiguity can be eliminated by using the unique lexico minimal summary vectors $x(s)$. However, there remains a significant ambiguity in computing D_q, since Definition 9.3 refers to a range of s values over which approximate equality holds. Let this range be denoted by $[L, U]$, where $L < U$. It is well known that, in general, the numerical value of any fractal dimension depends on the range of box sizes over which the dimension is computed. What had not been previously recognized is that for a complex network the choice of L and U can dramatically change the shape of the D_q vs. q curve: depending on L and U, 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 [49]. Example 9.1 and Fig. 9.1 provided an example where the D_q vs. q plot is not monotone non-increasing, even for the simple case $[L, U] = [2, 3]$. This behavior stands in sharp contrast to the behavior of a geometric multifractal, for which it is known [20] that D_q is non-increasing in q.

Recalling that $\log Z\big(x(s), q\big)$ for a complex network $\mathbb{G}$ is defined by (9.4), one way to compute D_q for a given q is to determine a range $[L_q, U_q]$ of s over which $\log Z\big(x(s), q\big)$ is approximately linear in $\log s$, and then use (9.5) to estimate D_q, e.g., using linear regression. With this approach, to report computational results to other researchers, it would be necessary to specify, for each q, the range of box sizes used to estimate D_q. This is certainly not the protocol currently followed in research on generalized dimensions. Rather, the approach taken in [49] and [67] is to pick a single L and U and estimate D_q for all q with this L and U. Moreover, rather than estimating D_q using a technique such as regression over the range $[L, U]$ of box sizes, [49] instead estimates D_q using only the two box sizes L and U. (As discussed in Chap. 5, such a two-point estimate was also used in [46], where it was shown that even for as simple a network as a one-dimensional chain, estimates of d_C obtained from regression do not behave well, and a two-point estimate has very desirable properties.)

E. Rosenberg, *A Survey of Fractal Dimensions of Networks*, SpringerBriefs in Computer Science, https://doi.org/10.1007/978-3-319-90047-6_10

With this two-point approach, the estimate of D_q is $1/(q-1)$ times the slope of the secant line connecting the points

$$\big(\log L, \log Z\big(x(L), q\big)\big) \quad \text{and} \quad \big(\log U, \log Z\big(x(U), q\big)\big) \ ,$$

where $x(L)$ and $x(U)$ are the lexico minimal summary vectors for box sizes L and U, respectively. Using (9.4) and (9.5), this secant estimate of D_q, which we denote by $D_q(L, U)$, is defined by

$$\begin{aligned} D_q(L, U) &\equiv \frac{\log Z\big(x(U), q\big) - \log Z\big(x(L), q\big)}{(q-1)\big(\log(U/\Delta) - \log(L/\Delta)\big)} \\ &= \frac{1}{(q-1)\log(U/L)} \log\left(\frac{\sum_{B_j \in \mathscr{B}(U)}[x_j(U)]^q}{\sum_{B_j \in \mathscr{B}(L)}[x_j(L)]^q}\right) . \end{aligned} \tag{10.1}$$

Example 10.1 Figure 10.1 plots box counting results for the *dolphins* network, which has 62 nodes, 159 arcs, and $\Delta = 8$. This is a social network describing frequent associations between 62 dolphins in a community living off Doubtful Sound, New Zealand [35]. For this network, and for all other networks described in this chapter, each lexico minimal summary vector $x(s)$ was computed using Procedure 9.1 and the graph coloring heuristic described in [48]. Figure 10.1 shows that the $\big(-\log(s/\Delta), \log B(s)\big)$ curve is approximately linear for $2 \le s \le 6$.

Figure 10.2 plots $\log Z\big(x(s), q\big)$ vs. $\log(s/\Delta)$ for $2 \le s \le 6$ and for $q = 2, 4, 6, 8, 10$ ($q = 2$ is the top curve, and $q = 10$ is the bottom curve). Figure 10.2 shows that, although the $\log Z\big(x(s), q\big)$ vs. $\log(s/\Delta)$ curves are less linear as q increases, a linear approximation is quite reasonable. Moreover, we are particularly interested in the behavior of the $\log Z\big(x(s), q\big)$ vs. $\log(s/\Delta)$ curve for small positive q, the region where the linear approximation is best. Using (10.1), Fig. 10.3 plots the secant estimate $D_q(L, U)$ vs. q for various choices of L and U. Since the D_q vs. q

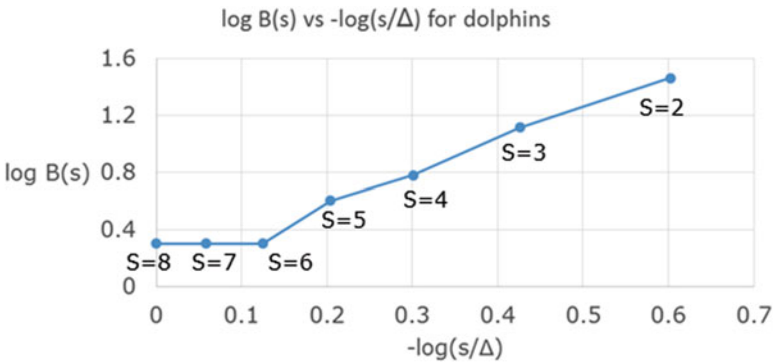

Fig. 10.1 Box counting for the *dolphins* network

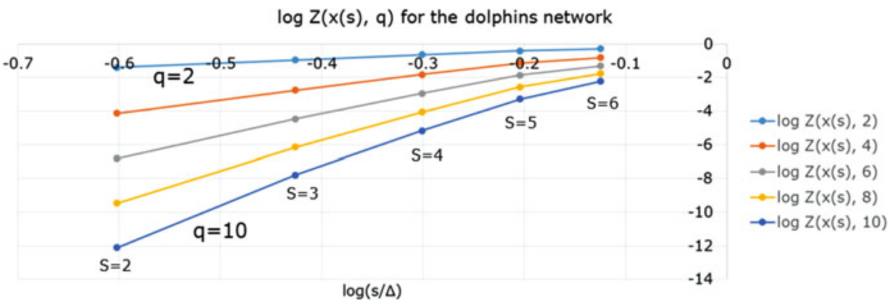

Fig. 10.2 $\log Z\big(x(s), q\big)$ vs. $\log(s/\Delta)$ for the *dolphins* network

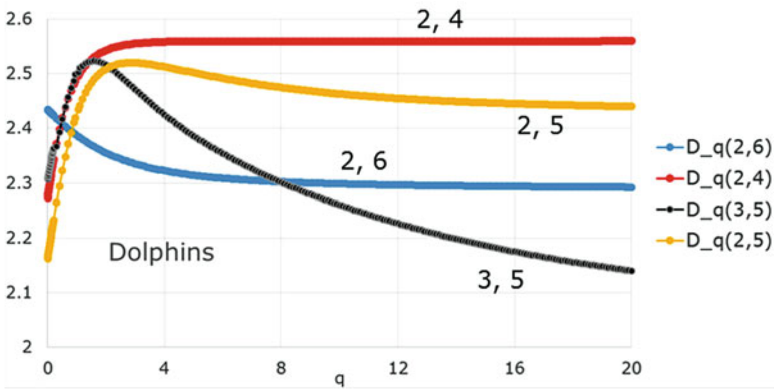

Fig. 10.3 Secant estimate of D_q for the *dolphins* network for different (L, U)

curve for a geometric multifractal is monotone non-increasing, it is remarkable that different choices of L and U lead to such different shapes for the $D_q(L, U)$ vs. q curve for the *dolphins* network. □

Let $D_0'(L, U)$ denote the first derivative with respect to q of the secant $D_q(L, U)$, evaluated at $q = 0$. A simple closed-form expression for $D_0'(L, U)$ is derived in [49]. For box size s, let $x(s) = \sum \mathscr{B}(s)$ be lexico minimal. Define

$$G(s) \equiv \left(\prod_{j=1}^{B(s)} x_j(s) \right)^{1/B(s)}$$

$$A(s) \equiv \frac{1}{B(s)} \sum_{j=1}^{B(s)} x_j(s)$$

$$R(s) \equiv \frac{G(s)}{A(s)} \tag{10.2}$$

so $G(s)$ is the geometric mean of the box masses summarized by $x(s)$, $A(s)$ is the arithmetic mean of the box masses summarized by $x(s)$, and $R(s)$ is the ratio of the geometric mean to the arithmetic mean. By the classic arithmetic-geometric inequality, for each s we have $R(s) \le 1$. Since $\sum_{j=1}^{B(s)} x_j(s) = N$, then $B(s)\,A(s) = N$. Theorems 10.1 and 10.2 below are proved in [49].

Theorem 10.1

$$D_0'(L, U) = \frac{1}{\log(U/L)} \log \frac{R(L)}{R(U)} \,.$$

□

Theorem 10.1 says that the slope of the secant estimate of D_q at $q = 0$ depends on $x(L)$ and $x(U)$ only through the ratio of the geometric mean to the arithmetic mean of the components of $x(L)$, and similarly for $x(U)$. Since $L < U$, Theorem 10.1 immediately implies the following corollary.

Corollary 10.1 *$D_0'(L, U) > 0$ if and only if $R(L) > R(U)$, and $D_0'(L, U) < 0$ if and only if $R(L) < R(U)$.* □

For a given L and U, Theorem 10.2 below provides a sufficient condition for $D_q(L, U)$ to have a local maximum or minimum.

Theorem 10.2 *(i) If $R(L) > R(U)$ and*

$$\frac{B(L)}{B(U)} > \frac{x_1(U)}{x_1(L)}$$

then $D_q(L, U)$ has a local maximum at some $q > 0$. (ii) If $R(L) < R(U)$ and

$$\frac{B(L)}{B(U)} < \frac{x_1(U)}{x_1(L)}$$

then $D_q(L, U)$ has a local minimum at some $q > 0$. □

Example 10.2 To illustrate Theorem 10.2, consider the *dolphins* network of Example 10.1 with $L = 3$ and $U = 5$. We have $B(3) = 13$ and $B(5) = 4$, so $D_0 = \log(13/4)/\log(5/3) \approx 2.307$. Also, $x_1(3) = 10$ and $x_1(5) = 28$, so by (9.8) we have $D_\infty \approx \log(28/10)/\log(5/3) \approx 2.106$. We have $R(3) \approx 0.773$, $R(5) \approx 0.660$, and $D_0'(L, U) \approx 0.311$. Hence $D_q(3, 5)$ has a local maximum, as seen in Fig. 10.3. Moreover, for the *dolphins* network, choosing $L = 2$ and $U = 5$ we have $D_0 = \log(29/4)/\log(5/2) \approx 2.16$, and $D_\infty \approx \log(28/3)/\log(5/2) \approx 2.44$, so $D_0 < D_\infty$, as is evident from Fig. 10.3. Thus the inequality $D_0 \ge D_\infty$, which is valid for geometric multifractals, does not hold for the *dolphins* network with $L = 2$ and $U = 5$. □

If for $s = L$ and $s = U$ we can compute a minimal s-covering with equal box masses, then $\mathbb{G}$ is a monofractal but not a multifractal. To see this, suppose all boxes in $\mathscr{B}(L)$ have the same mass, and that all boxes in $\mathscr{B}(U)$ have the same mass. Then for $s = L$ and $s = U$ we have $x_j(s) = N/B(s)$ for $1 \leq j \leq B(s)$, and (9.1) yields

$$Z\big(x(s), q\big) = \sum_{B_j \in \mathscr{B}(s)} \left(\frac{x_j(s)}{N}\right)^q = \sum_{B_j \in \mathscr{B}(s)} \left(\frac{1}{B(s)}\right)^q = [B(s)]^{1-q} .$$

From (9.5), for $q \neq 1$ we have

$$\begin{aligned} D_q &= \frac{\log Z\big((x(U), q\big) - \log Z\big(x(L), q\big)}{(q-1)\big(\log U - \log L\big)} = \frac{\log\big([B(U)]^{1-q}\big) - \log\big([B(L)]^{1-q}\big)}{(q-1)\big(\log U - \log L\big)} \\ &= \frac{\log B(L) - \log B(U)}{\log U - \log L} = D_0 = d_B , \end{aligned} \tag{10.3}$$

so $\mathbb{G}$ is a monofractal. Thus equal box masses imply $\mathbb{G}$ is a monofractal, the simplest of all fractal structures.

There are several ways to try to obtain equal box masses in a minimal s-covering of $\mathbb{G}$. As discussed in Chap. 8, ambiguity in the choice of minimal coverings used to compute d_I is eliminated by maximizing entropy. Since the entropy of a probability distribution is maximized when all the probabilities are equal, a maximal entropy minimal covering equalizes (to the extent possible) the box masses. Similarly, as discussed in Chap. 9, ambiguity in the choice of minimal s-coverings used to compute D_q is eliminated by minimizing the partition function $Z_q\big(\mathscr{B}(s)\big)$. Since for all sufficiently large q the lexico minimal vector $x(s)$ summarizes the s-covering that minimizes $Z_q\big(\mathscr{B}(s)\big)$, and since for $q > 1$ a partition function is minimized when all the probabilities are equal, then $x(s)$ also equalizes (to the extent possible) the box masses. Theorem 10.1 suggests a third way to try to equalize the masses of all boxes in a minimal s-covering: since $G(s) \leq A(s)$ and $G(s) = A(s)$ when all boxes have the same mass, a minimal s-covering that maximizes $G(s)$ will also equalize (to the extent possible) the box masses. The advantage of computing the lexico minimal summary vectors $x(s)$, rather than maximizing the entropy or maximizing $G(s)$, is that, by Theorem 9.1, the summary vector $x(s)$ is unique.

We now apply Theorem 10.1 to the *chair* network, to the *dolphins* network, and to a *jazz* network.

Example 10.3 For the *chair* network of Fig. 8.2 we have $L = 2$, $x(L) = (2, 2, 1)$, $U = 3$, and $x(U) = (3, 2)$. We have $D_0'(2, 3) \approx -0.070$, as shown in Fig. 9.1 by the slightly negative slope of the lower curve at $q = 0$. As mentioned above, this curve is not monotone non-increasing; it has a local minimum. □

Example 10.4 For the *dolphins* network studied in Example 10.1, Table 10.1 provides $D_0'(L, U)$ for various choices of L and U. The values in Table 10.1 are

Table 10.1 $D_0'(L, U)$ for the *dolphins* network

L, U	$D_0'(L, U)$
2, 6	−0.056
3, 5	0.311
2, 4	0.393
2, 5	0.367

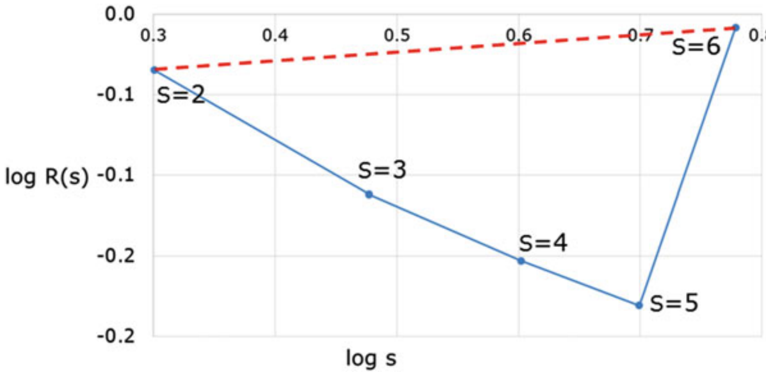

Fig. 10.4 log $R(s)$ vs. log s for the *dolphins* network

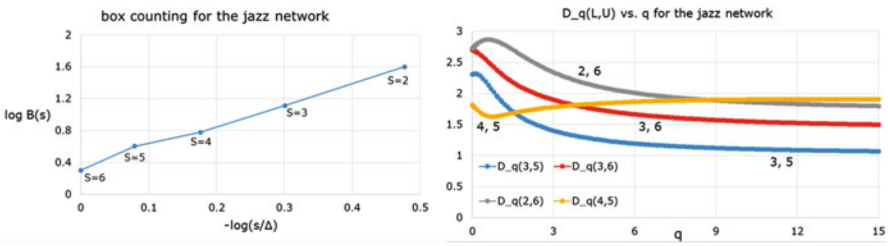

Fig. 10.5 *Jazz* box counting (left) and D_q vs. q for various L and U (right)

better understood using Fig. 10.4, which plots $\log R(s)$ vs. $\log s$. For example, for $(L, U) = (2, 6)$ we have $D_0'(2, 6) = \log\big(R(2)/R(6)\big)/\big(\log 6/2\big) \approx -0.056$, as illustrated by the slightly *positive* slope of the dashed red line in Fig. 10.4, since the slope of the dashed red line is $-D_0'(2, 6)$. For the other choices of (L, U) in Table 10.1, the values of $D_0'(L, U)$ are positive and roughly equal. Figure 10.2 visually suggests that $\log Z\big(x(s), q\big)$ is better approximated by a linear fit over $s \in [2, 5]$ than over $s \in [2, 6]$, and Fig. 10.4 clearly shows that $s = 6$ is an outlier in that using $U = 6$ dramatically changes $D_0'(L, U)$. □

Example 10.5 This network, with 198 nodes, 2742 arcs, and diameter 6, is a collaboration network of jazz musicians [19]. Figure 10.5 shows the results of box counting; the curve appears reasonably linear for $s \in [2, 6]$. Figure 10.5 also plots $D_q(L, U)$ vs. q for four choices of L and U. Table 10.2 provides $D_0'(L, U)$, D_0, and D_∞ for nine choices of L and U; the rows are sorted by decreasing $D_0'(L, U)$.

Table 10.2 Results for the *jazz* network for various L and U

L, U	$D'_0(L, U)$	D_0	D_∞
2, 3	1.576	2.77	2.16
2, 4	1.224	2.74	1.42
2, 5	0.826	2.51	1.51
3, 4	0.728	2.69	0.37
2, 6	0.485	2.73	1.68
3, 5	0.231	2.31	1.00
3, 6	−0.154	2.70	1.40
4, 5	−0.411	1.82	1.82
5, 6	−1.232	3.80	2.52

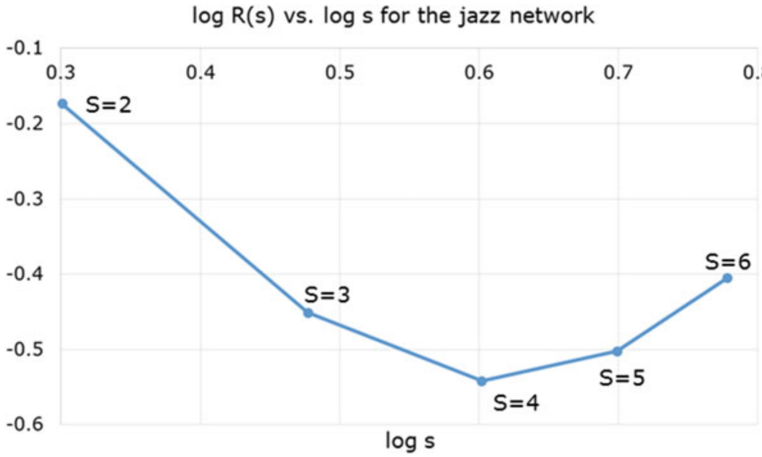

Fig. 10.6 $\log R(s)$ vs. $\log s$ for the *jazz* network

It is even possible for the $D_q(L, U)$ vs. q curve to exhibit both a local maximum and a local minimum: for the *jazz* network with $L = 4$ and $U = 5$, there is a local minimum at $q \approx 0.7$ and a local maximum at $q \approx 12.8$. Figure 10.6 plots $\log R(s)$ vs. $\log s$ for the *jazz* network. □

These results, together with the results in [47, 48], show that two requirements should be met when reporting fractal dimensions of a complex network. First, since there are in general multiple minimal s-coverings, and these different coverings can yield different values of D_q, computational results should specify the rule (e.g., a maximal entropy covering, or a covering yielding a lexico minimal summary vector) used to unambiguously select a minimal s-covering. Second, the lower bound L and upper bound U on the box sizes used to compute D_q should be reported. Published values of D_q not meeting these two requirements cannot in general be considered benchmarks. As to the values of L and U yielding the most meaningful results, it is desirable to identify the largest range $[L, U]$ over which $\log Z$ is approximately linear in $\log s$; this is a well-known principle in the estimation of fractal dimensions. Future research may uncover, based on the $\log R(s)$ vs. $\log s$ curve, other criteria for selecting L and U.