Chapter 2
Covering a Complex Network

The previous chapter showed how the box counting and Hausdorff dimensions of a geometric object Ω are computed from a covering of Ω. With this background, we can now consider what it means to cover a complex network $\mathbb{G}$, and how a fractal dimension can be computed from a covering of $\mathbb{G}$. We require some definitions. The network B is a *subnetwork* of $\mathbb{G}$ if B can be obtained from $\mathbb{G}$ by deleting nodes and arcs. By a *box* we mean a subnetwork of $\mathbb{G}$. A box is *disconnected* if some nodes in the box cannot be connected by arcs in the box. Let $\{B_j\}_{j=1}^{J} \equiv \{B_1, B_2, \cdots, B_J\}$ be a collection of boxes. Two types of coverings of $\mathbb{G}$ have been proposed: *node coverings* and *arc coverings*. Let s be a positive integer.

Definition 2.1 (*i*) The set $\{B_j\}_{j=1}^{J}$ is a *node* s-covering of $\mathbb{G}$ if for each j we have $diam(B_j) < s$ and if each node in $\mathbb{N}$ is contained in exactly one B_j. (*ii*) The set $\{B_j\}_{j=1}^{J}$ is an *arc* s-covering of $\mathbb{G}$ if for each j we have $diam(B_j) < s$ and if each arc in $\mathbb{A}$ is contained in exactly one B_j. □

If B_j is a box in a node or arc s-covering of $\mathbb{G}$ then the requirement $diam(B_j) < s$ in Definition 2.1 implies that B_j is connected. However, this requirement, which is a standard assumption in defining the box counting dimension of $\mathbb{G}$ (e.g., [16, 29, 30, 48, 56]), may frequently be violated, for good reasons, in some methods for determining the fractal dimensions of $\mathbb{G}$, as we will discuss in Sect. 3.6.

It is possible to define a node covering of $\mathbb{G}$ to allow a node to be contained in more than one box; coverings with possibly overlapping boxes are studied in [15, 60]. The great advantage of non-overlapping boxes is that they immediately yield a probability distribution, as discussed in Chap. 8. The probability distribution obtained from a non-overlapping node covering of $\mathbb{G}$ is the basis for computing the information dimension d_I and the generalized dimensions D_q of $\mathbb{G}$ (Chap. 9). Therefore, in this survey, each node covering of $\mathbb{G}$ is assumed to use non-overlapping boxes, as specified in Definition 2.1.

E. Rosenberg, *A Survey of Fractal Dimensions of Networks*, SpringerBriefs in Computer Science, https://doi.org/10.1007/978-3-319-90047-6_2

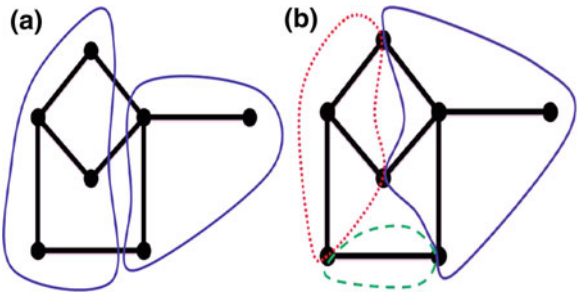

Fig. 2.1 A node 3-covering (**a**) and an arc 3-covering (**b**)

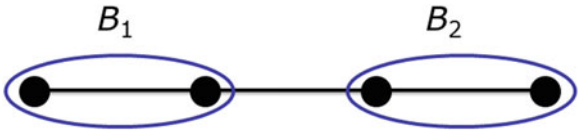

Fig. 2.2 A node covering which is not an arc covering

Since a network of diameter 0 contains only a single node, a node 1-covering contains N boxes. Since a node 1-covering provides no useful information other than N itself, we consider node s-coverings only for $s \geq 2$. Figure 2.1a illustrates a node 3-covering with $J = 2$. Both boxes in the 3-covering have diameter 2. Figure 2.1b illustrates an arc 3-covering of the same network using three boxes. The box indicated by the solid blue line contains four arcs, the box indicated by the dotted red line contains three arcs, and the box indicated by the dashed green line contains one arc.

Each arc covering of $\mathbb{G}$ yields a node covering. However, the converse is not true: a node covering does not in general yield an arc covering. This is illustrated by the simple example of Fig. 2.2. The nodes are covered by B_1 and B_2, but the arc in the middle belongs to neither B_1 nor B_2.

Definition 2.2 (*i*) An arc s-covering $\{B_j\}_{j=1}^{J}$ is *minimal* if for any other arc s-covering $\{B_j'\}_{j=1}^{J'}$ we have $J \leq J'$. (*ii*) A node s-covering $\{B_j\}_{j=1}^{J}$ is *minimal* if for any other node s-covering $\{B_j'\}_{j=1}^{J'}$ we have $J \leq J'$. □

That is, a covering is minimal if it uses the fewest possible number of boxes. For $s > \Delta$, the minimal node or arc s-covering consists of a single box, which is $\mathbb{G}$ itself. Virtually all research has considered node coverings; only a few studies (e.g., [74]) use arc coverings. The reason arc coverings are rarely used is that, in practice, computing a fractal dimension of a geometric object typically starts with a given set of points in $\mathbb{R}^E$ (the points are then covered by boxes, or the distance between each pair of points is computed (e.g., [22, 36])), and nodes in a network are analogous to points in $\mathbb{R}^E$. Having very briefly contrasted arc coverings and node coverings for a network, we now abandon arc coverings; henceforth, all coverings of $\mathbb{G}$ are node coverings, and by covering $\mathbb{G}$ we mean covering the nodes of $\mathbb{G}$. Also, henceforth by an s-covering we mean a node s-covering, and by a covering of size s we mean an s-covering.

2.1 Box Counting with Diameter-Based or Radius-Based Boxes

There are two main approaches used to define boxes for use in covering $\mathbb{G}$: diameter-based boxes and radius-based boxes.

Definition 2.3 (*i*) A radius-based box $\mathbb{G}(n, r)$ with center node $n \in \mathbb{N}$ and radius r is the subnetwork of $\mathbb{G}$ containing all nodes whose distance to n does not exceed r. Let $B_R(r)$ be the minimal number of radius-based boxes of radius at most r needed to cover $\mathbb{G}$. (*ii*) A diameter-based box $\mathbb{G}(s)$ of size s is a subnetwork of $\mathbb{G}$ of diameter $s-1$. Let $B_D(s)$ denote the minimal number of diameter-based boxes of size at most s needed to cover $\mathbb{G}$. □

Thus the node set of $\mathbb{G}(n, r)$ is $\{x \in \mathbb{N} \mid dist(n, x) \leq r\}$. Radius-based boxes are used in the *Maximum Excluded Mass Burning* and *Random Sequential Node Burning* methods described in Chap. 3. Interestingly, the above definition of a radius-based box may frequently be violated in the *Maximum Excluded Mass Burning* and *Random Sequential Node Burning* methods. In particular, some radius-based boxes created by those methods may be disconnected, or some boxes may contain only some of the nodes whose distance to the center node n does not exceed r.

A diameter-based box $\mathbb{G}(s)$ is not defined in terms of a center node; instead, for $x, y \in \mathbb{G}(s)$ we require $dist(x, y) < s$. Diameter-based boxes are used in the *Box Burning* and *Compact Box Burning* heuristics described in Chap. 3. The above definition of a diameter-based box also may frequently be violated in the *Box Burning* and *Compact Box Burning* methods. Also, since each node in $\mathbb{G}$ must belong to exactly one B_j in an s-covering $\{B_j\}_{j=1}^{J}$ using diameter-based boxes, then in general we will not have $diam(B_j) = s - 1$ for all j. To see this, consider a chain of three nodes (call them x, y, and z), and let $s = 2$. The minimal 2-covering using diameter-based boxes requires two boxes, B_1 and B_2. If B_1 covers x and y then B_2 covers only z, so the diameter of B_2 is 0.

The minimal number of diameter-based boxes of size at most $2r + 1$ needed to cover $\mathbb{G}$ is, by definition, $B_D(2r + 1)$. We have $B_D(2r + 1) \leq B_R(r)$ [29]. To see this, let $\mathbb{G}(n_j, r_j)$, $j = 1, 2, \cdots, B_R(r)$ be the boxes in a minimal covering of $\mathbb{G}$ using radius-based boxes of radius at most r. Then $r_j \leq r$ for all j. Pick any j, and consider box $\mathbb{G}(n_j, r_j)$. For any nodes x and y in $\mathbb{G}(n_j, r_j)$ we have

$$dist(x, y) \leq dist(x, n_j) + dist(n_j, y) \leq 2r_j \leq 2r\,,$$

so $\mathbb{G}(n_j, r_j)$ has diameter at most $2r$. Thus these $B_R(r)$ boxes also serve as a covering of size $2r + 1$ using diameter-based boxes. Therefore, the minimal number of diameter-based boxes of size at most $2r+1$ needed to cover $\mathbb{G}$ cannot exceed $B_R(r)$; that is, $B_D(2r + 1) \leq B_R(r)$.

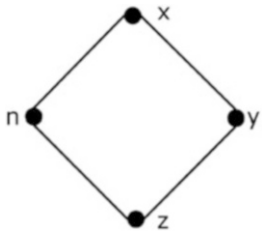

Fig. 2.3 Diameter-based vs. radius-based boxes

The reverse inequality does not in general hold, since a diameter-based box of size $2r+1$ can contain more nodes than a radius-based box of radius r. For example, consider the network $\mathbb{G}$ of Fig. 2.3. The only nodes adjacent to n are x and z, so $\mathbb{G}(n, 1) = \{n, x, z\}$ and $B_R(1) = 2$. Yet the diameter of $\mathbb{G}$ is 2, so it can be covered by a single diameter-based box of size 3, namely $\mathbb{G}$ itself, so $B_D(3) = 1$. Thus $B_R(r)$ and $B_D(2r+1)$ are not in general equal. Nonetheless, for the *C. elegans* and Internet backbone networks studied in [56], the calculated fractal dimension was the same whether radius-based or diameter-based boxes were used. Similarly, both radius-based and diameter-based boxes yielded a fractal dimension of approximately 4.1 for the WWW (the World Wide Web) [29].

The term *box counting* refers to computing a minimal s-covering of $\mathbb{G}$ for a range of values of s, using either radius-based boxes or diameter-based boxes. Conceivably, other types of boxes might be used to cover $\mathbb{G}$. In the fractal literature, the box counting dimension d_B is often informally defined by the scaling $B_D(s) \sim s^{-d_B}$. (The symbol "$\sim$", frequently used in the fractal literature but often with different meanings, should here be interpreted to mean "approximately behaves like".) Definition 2.4 below provides a more computationally useful definition of d_B for a complex network.

Definition 2.4 $\mathbb{G}$ has box counting dimension d_B if over some range of s and for some constant c we have

$$\log B_D(s) \approx -d_B \log(s/\Delta) + c\,. \quad \square \tag{2.1}$$

Alternatively, (2.1) can be written as $\log B_D(s) \approx -d_B \log s + c$. If $\mathbb{G}$ has box counting dimension d_B then over some range of s we have $B_D(s) \approx a s^{-d_B}$ for some constant a. In the terminology of [16], if the box counting dimension for $\mathbb{G}$ exists, then $\mathbb{G}$ enjoys the *fractal scaling property*, or, more simply, $\mathbb{G}$ is *fractal*. The main feature apparently displayed by fractal networks is a repulsion between hubs, where a hub is a node with a significantly higher node degree than a non-hub node. That is, the highly connected nodes tend to be not directly connected [72]. This tendency can be quantified using the joint node degree distribution $p(\delta_1, \delta_2)$ that a node with degree δ_1 and a node with degree δ_2 are neighbors (i.e., connected by a single arc). In contrast, for a non-fractal network $\mathbb{G}$, hubs are mostly connected to other hubs, which implies that $\mathbb{G}$ enjoys the small-world property [16]. (Roughly speaking, $\mathbb{G}$ is a *small-world* network if $diam(G)$ grows as $\log(N)$ [57].) Also, the concepts of

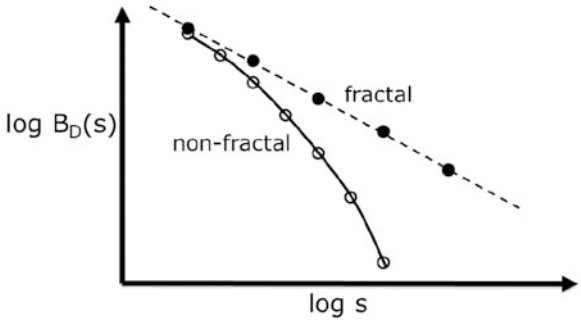

Fig. 2.4 Fractal vs. non-fractal scaling

modularity and *fractality* for a network are closely related. Interconnections within a module (e.g., a biological sub-system) are more prevalent than interconnections between modules. Similarly, in a fractal network, interconnections between a hub and non-hub nodes are more prevalent than interconnections between hubs. Non-fractal networks are typically characterized by a sharp decay of $B_D(s)$ with s, which is better described by an exponential law $B_D(s) \sim e^{-\beta s}$, where $\beta > 0$, rather than by a power law $B_D(s) \sim s^{-\beta}$, with a similar statement holding if radius-based boxes are used. These two cases are illustrated in Fig. 2.4, taken from [16], where the solid circles are measurements from a fractal network, and the hollow circles are from a non-fractal network.