Journal of Statistical Physics, Vol. 64, Nos. 1/2, 1991

Calculating the Hausdorff Dimension of Tree Structures

Lorraine M. Rellick,[1,2] Gerald A. Edgar,[3] and Michael H. Klapper[1]

Received October 22, 1990; final February 26, 1991

The dimension of certain tree structures is of importance in percolation theory, as well as in the theoretical treatment of many other branching processes. We present a method of determining the Hausdorff dimension of such structures by employing the technique of Mauldin and Williams. The dimension is calculated based on the probability of generation of each branch from its parent on the tree representing the process. We use this method to analyze the dimension of tree structures representing two-directional linear bonding between equally weighted monomers, and show how it can be used to model enzymatic reaction pathways.

KEY WORDS: Branching process; enzyme reaction kinetics; Hausdorff dimension; Mauldin–Williams graph; percolation theory; tree representations.

1. INTRODUCTION

Trees have been used to model many different physical processes. These include the modeling of percolation clusters and the growth of aggregates,[1] the hierarchical treatment of energy barriers in molecular systems,[2] and the conformational steps followed by a macromolecule as it folds from an extended to a globular structure.[3] The dimension of the resulting structure has been shown to be important in modeling random walks, diffusion, and kinetic and thermodynamic functions. A recent method for determining the Hausdorff dimension developed by Mauldin and Williams can be applied to such tree representations.

[1] Department of Chemistry, Ohio State University, Columbus, Ohio 43210.
[2] Current address: Department of Biochemistry, Ohio State University, Columbus, Ohio 43210.
[3] Department of Mathematics, Ohio State University, Columbus, Ohio 43210.

0022-4715/91/0700-0077$06.50/0

2. METHOD[4,5]

The Mauldin–Williams formalism is based on the concept of graph self-similarity. In general this is a treatment for systems which can be visualized as a group (set) of sets. Each of these sets can be described as being constructed from versions (whole or shrunken, rotated, translated, or a combination of these) of some or all of the sets in the group, including itself. A similarity factor or ratio is associated with each of these versions which construct a set in the group. The Mauldin–Williams representation of such a system is a directed multigraph. The nodes or vertices of the graph represent the sets in the group, and the directed edges represent how a particular node or set is composed from the others (and/or itself). The edges are directed from the constructed set to those from which it is constructed. Another way of looking at it is that the edges are directed to those sets into which the set can be decomposed. The similarity factor associated with a particular edge is written along side it. Figure 1 shows an example of a Mauldin–Williams graph representing a geometric system.

Tree structures are analogous to the system of sets described above. The edges in this case would correspond to the children into which the parent is "decomposed." The similarity ratio can correspond to the probability of generating that child from the parent, or with some other factor associated with generation of that element of the tree.

To find the dimension of the Mauldin–Williams representation, it is in general necessary to first compute Perron numbers for the system. These are positive numbers which obey the following relation:

$$q_i^s = \sum_{\substack{j \in V \\ e \in \mathscr{E}_{i,j}}} r^s(e)\, q_j^s \tag{2.1}$$

where $r(e)$ is the similarity ratio associated with edge e, V refers to the set of vertices, and $\mathscr{E}$ to the set of edges ($\mathscr{E}_{i,j}$ is the set of edges directed from

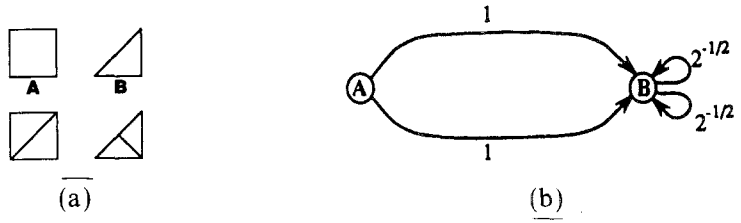


Fig. 1. Example of a system showing graph self-similarity. (a) This system is composed of a square and a triangle. The relationship between the two components of the system can be stated as: A = two full-size copies of B, and B = two $1/\sqrt{2}$-size copies of B. (b) The Mauldin–Williams graph for this system. Notice that it is not strongly connected, and not contracting [not all $r(e)$'s are less than 1].

i to j). The positive number s will be unique and is the dimension of the graph. The Perron numbers can be calculated analytically or numerically, depending on the system. However, if there is only one node in the graph, the relation reduces to

$$1 = \sum_{e \in \mathscr{E}} r^s(e) \tag{2.2}$$

In this case, dimension s is simply the solution to this equation.

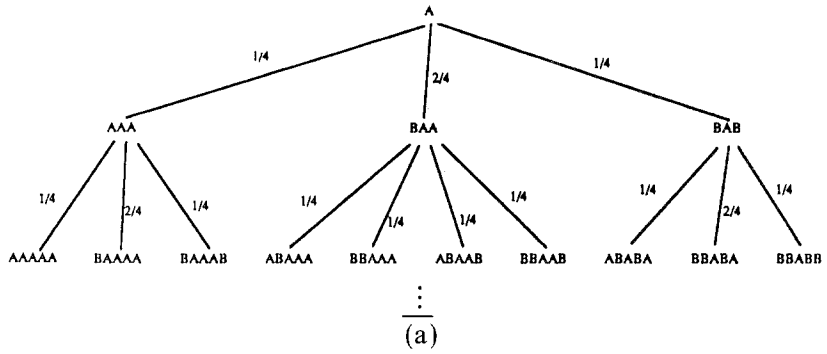
If the graph is strongly connected (there is a path from every node to every other node on the graph), then the sets (the vertices or nodes of the graph) will all have the same dimension, and this will be labeled the dimension of the graph. If the graph is not strongly connected, it may be possible to decompose it into strongly connected subgraphs. In this case, the dimension of the system will be the maximum dimension of those found for the decomposition subgraphs (see ref. 5, Theorem 4). If the system of sets is strongly connected (with the exception above), and contracting (i.e., all of the similarity ratios are less than one), and an open set condition is satisfied for the realization of the graph,[4] then the dimension s of the graph will be the Hausdorff dimension of the system.[4]

3. RESULTS

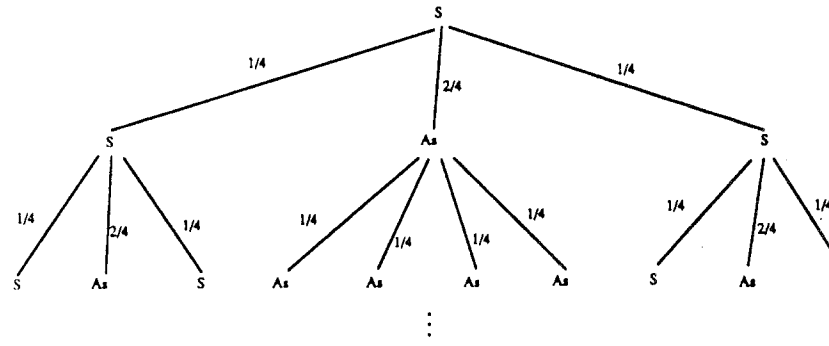
Here we give examples of the calculation of the Hausdorff dimension for tree structures, two-directional linear bonding between equally weighted monomers, and the reaction pathway of a simple enzyme-catalyzed reaction.

Consider a linear polymer chain which is composed of two monomers, A and B. These monomers can add to the chain in either direction, and the addition of either an A or a B is equally likely. In other words, the probability of either an A or a B bonding to the chain is the same, 0.5. Because the monomers can bind to either end of the chain, the probability of an A bonding on both ends of the existing chain is $(0.5 \times 0.5) = 0.25$. The probability of a B joining on both ends will also be 0.25. However, the probability of a species occurring with an A bonding on one end and a B bonding on the other is more interesting because in some cases the parent will be a symmetric species (about the center node), and in others, asymmetric. If the parent is asymmetric, two distinguishable species will be formed (Fig. 2), each having probability 0.25 of being generated from the

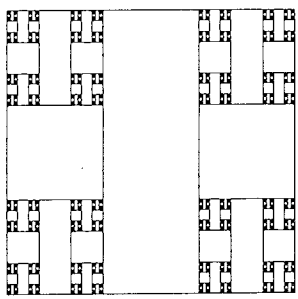
[4] For details about the open set condition, see refs. 4 and 5. Satisfying this condition assures no overlap between the sets that compose the system. In the case of a tree structure, no overlap is possible and so for this specific case, the condition is not required.



(a)



(b)



(c)

Fig. 2. Two-dimensional linear bonding between equally-weighted monomers. (a) Tree representation. *A* and *B* are equally-weighted monomers. The probability of a branch occurring is shown next to the branch. (b) Tree representation. *S* and *As* refer to symmetric and asymmetric species [for species, see tree in panel (a)]. (c) Box representation (due to G. A. Edgar) of the symmetric transition in this system.

parent. However, if the parent is a symmetric species, only one child node will be formed, having a probability of 0.5. This event is twice as likely because two identical species will be formed. These resulting species will be asymmetric about the center node. An interesting tree structure is generated from this process because of the symmetric/asymmetric transition.

A Mauldin–Williams graph for the tree structure is shown in Fig. 3. Notice that, because of the symmetric/asymmetric transition, the tree is not self-similar. In other words, the pattern of nodes on a certain tier in the tree is not in all cases exactly like the pattern of nodes on the tier immediately above. The resulting Mauldin–Williams representation will not be strongly connected. This graph, though, can be decomposed into two strongly connected subgraphs, and so the Hausdorff dimension can be evaluated. As stated above, Hausdorff dimension will be the maximum of the s values for the decomposition subgraphs.

In this example, the s value for the first decomposition graph (left graph in Fig. 3b) will be 0.5, and that for the second (right graph in Fig. 3b) is 1.0. The overall Hausdorff dimension for this tree is therefore 1.0.

Consider the tree structure shown in Fig. 4a. This structure represents the binding/reaction pathway for simple enzyme-catalyzed reactions. The Mauldin–Williams graph resulting from this tree will be slightly more complicated than in the previous case, and is shown in Fig. 4c.

This graph, unlike that of the previous example, is strongly connected, and has three nodes, E, ES, and EP, and five edges with ratios a, b, c, d, and e. The nodes represent the free enzyme, the enzyme–substrate complex,

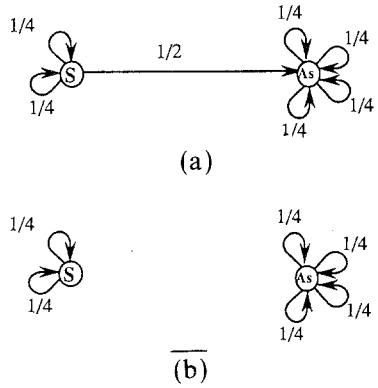


Fig. 3. (a) Mauldin–Williams graph of the tree shown in Fig. 2b. Notice that this graph is contracting, but not strongly connected. (b) Decomposition into strongly connected, contracting subgraphs.

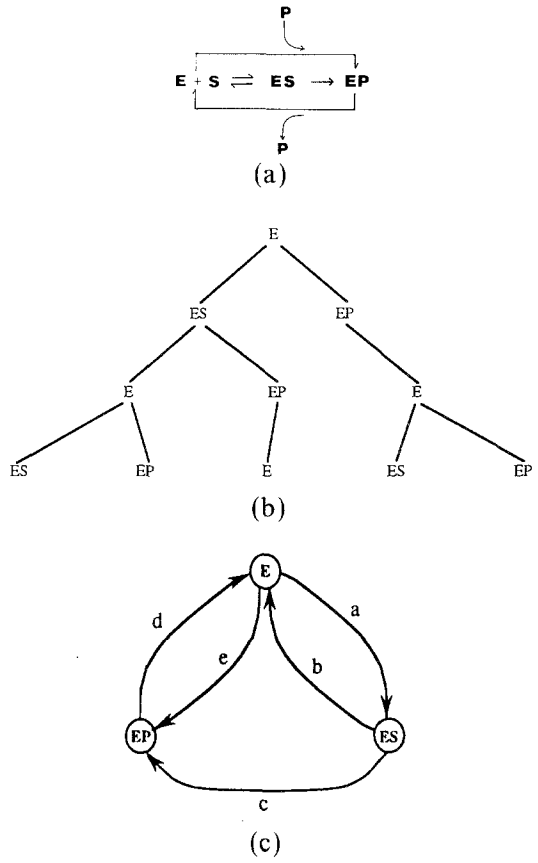


Fig. 4. A simple enzymatic binding/reaction pathway (a) Conventional representation, (b) tree representation, (c) Mauldin–Williams graph for this system. The lower case letters represent the probabilities associated with each step.

and the enzyme–product complex. The values associated with the edges represent the probability of the binding or reaction along the direction of the arrow.

The expressions for the Perron numbers of this graph allow us to determine the dimension of the system. As would be expected, this dimension will change with the edge (ratio) values a, b, c, d, and e. Using (2.1), and following the procedure given in ref. 5, we obtain

$$V = \{E, ES, EP\}$$

$$q_E^s = \sum_{\substack{j \in V \\ e \in \mathcal{E}_{E,j}}} r^s(e)\, q_j^s = a^s q_{ES}^s + e^s q_{EP}^s \tag{3.1a}$$

$$q_{ES}^s = \sum_{\substack{j \in V \\ e \in \mathscr{E}_{ES,j}}} r^s(e)\, q_j^s = b^s q_E^s + c^s q_{EP}^s \tag{3.2a}$$

$$q_{EP}^s = \sum_{\substack{j \in V \\ e \in \mathscr{E}_{EP,j}}} r^s(e)\, q_j^s = d^s q_E^s \tag{3.3a}$$

Letting $x = q_E^s$, $y = q_{ES}^s$, and $z = q_{EP}^s$, these become

$$x = a^s y + e^s z \tag{3.1b}$$

$$y = b^s x + c^s z \tag{3.2b}$$

$$z = d^s x \tag{3.3b}$$

Solving this system simultaneously leads to

$$x = a^s(b^s x + c^s d^s x) + e^s d^s x$$

Division by x and rearrangment yields

$$0 = a^s b^s + a^s c^s d^s + e^s d^s - 1 \tag{3.5}$$

Using this expression, s, the dimension of the system, can easily be determined numerically for various values of the probabilities a, b, c, d, and e. These probabilities can be determined by detailed experimentation on the system in question, and will be different, depending on the particular enzyme and reaction conditions (concentration of enzyme and substrate, exit route for product, temperature, pH, etc.). Depending on the probabilities expressed by the system, the dimension for the reaction can be widely different. Two examples are given below.

In Fig. 5, two sets of probabilities are given for the branches on the binding/reaction pathway. The probabilities used here were not derived from experimental data, but were chosen to emphasize the results of the calculations. In the case of Fig. 5a, the probability of the enzyme–substrate complex reacting to yield free enzyme and product is 0.005, and that of the enzyme–substrate complex dissociating into free enzyme and substrate is 0.995. These values are reversed in the case of Fig. 5b. Interchanging these values has major consequences in terms of the kinetics of the enzyme, as well as for the dimension of the resulting graph. Using the first set of probabilities, expression (3.5) becomes

$$\begin{aligned} 0 = {} & (0.995)^s\,(0.995)^s + (0.995)^s\,(0.005)^s\,(0.875)^s \\ & + (0.005)^s\,(0.875)^s - 1 \end{aligned} \tag{3.6}$$

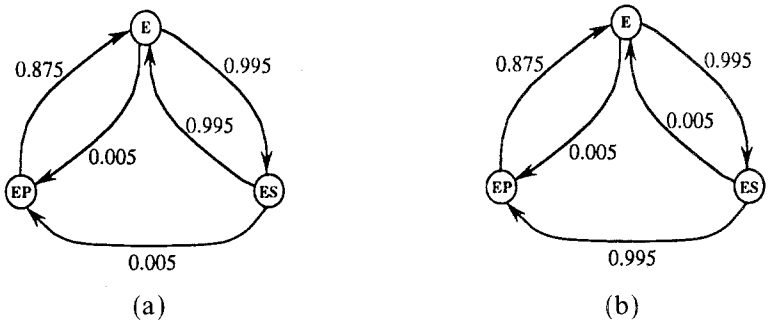


Fig. 5. Mauldin–Williams graphs for the simple enzyme system, showing two possible sets of probabilities. (a) Graph with dimension 0.9793, (b) graph with dimension 0.5954.

Letting $\lambda = (0.995)^s$, and expressing the other probabilities as powers of (0.995), this becomes

$$0 = \lambda^2 + \lambda^{1084.64} + \lambda^{1083.64} - 1 \tag{3.7a}$$

Solving (3.7a) iteratively yields a value of 0.9951032, or an s value of 0.9793. In the second case, expression (3.7) becomes

$$0 = \lambda^{1085} + \lambda^{28.64} + \lambda^{1083.64} - 1 \tag{3.7b}$$

Solving yields a value for λ of 0.9970199 and for s of 0.5954.

At this point we should note the following regarding these graphs. If, as in the first example, the sum of the edge ratios of the arrows leaving each of the nodes in the decomposition graph is equal to one, the dimension of that graph will equal one. In this example the sum of the ratios for the edges leaving node E is equal to 1, and the sum of those leaving ES is 1; however, the sum of those leaving node EP is 0.857.[5] Because not all of these values are equal to one, the dimension of this graph is a fractional value.

4. CONCLUSION

In this paper we have shown that a recently developed method for calculation of the Hausdorff dimension of systems showing graph self-similarity can be used to determine this dimension for systems which can

[5] In this example, mechanism based inhibition is occurring which effectively removes some of the enzyme from the pathway and which causes the sum of the probabilities associated with the arrows leaving the EP node to be less than one.

be modeled as tree structures. It is our hope that this may be of some use to persons interested in the dimension of the many physical processes which have been modeled using trees.

In addition, we have shown that this method can be used to determine the dimension of binding/reaction pathways. Futher study, which we are currently undertaking, is necessary to determine the thermodynamic and kinetic consequences of the differences in dimensionality of reactions studied in this manner. We feel that the dimension of a reaction can be related to its reaction mechanism, and we hope to show the nature of this relationship in a later publication.

REFERENCES

1. S. Havlin, J. E. Kiefer, F. Leyvraz, and G. H. Weiss, Probability distributions for percolation clusters generated on a Cayley tree at criticality, *J. Stat. Phys.* **47**:173 (1987).
2. T. A. Vilgis, On the dynamics of simple hierarchia, *J. Stat. Phys.* **47**:133 (1987).
3. T. F. Havel, The combinatorial distance geometry approach to the calculation of molecular conformation, Ph.D. Dissertation, University of California, Berkeley (1982).
4. G. A. Edgar, *Measure, Topology, and Fractal Geometry* (Springer-Verlag, New York, 1990).
5. R. D. Mauldin and S. C. Williams, Hausdorff dimension in graph directed constructions, *Trans. Am. Math. Soc.* **309**:811 (1988).