Fractal Networks: topology, dimension and complexity

L. Bunimovich[1] and P. Skums[2]
[1)]*School of Mathematics, Georgia Institute of Technology, 686 Cherry St NW, Atlanta, GA 30332*
[2)]*School of Computing, University of Connecticut, 371 Fairfield Way, Storrs, CT, 06269*

(*Electronic mail: pavel.skums@uconn.edu)

(*Electronic mail: leonid.bunimovich@math.gatech.edu)

(Dated: 22 September 2024)

Over the past two decades, the study of self-similarity and fractality in discrete structures, particularly complex networks, has gained momentum. This surge of interest is fueled by the theoretical developments within the theory of complex networks and the practical demands of real-world applications. Nonetheless, translating the principles of fractal geometry from the domain of general topology, dealing with continuous or infinite objects, to finite structures in a mathematically rigorous way poses a formidable challenge.

In this paper, we overview such theory that allows to identify and analyze fractal networks through the innate methodologies of graph theory and combinatorics. It establishes the direct graph-theoretical analogues of topological (Lebesgue) and fractal (Hausdorff) dimensions in a way that naturally links them to combinatorial parameters that have been studied within the realm of graph theory for decades. This allows to demonstrate that the self-similarity in networks is defined by the patterns of intersection among densely connected network communities. Moreover, the theory bridges discrete and continuous definitions by demonstrating how the combinatorial characterization of Lebesgue dimension via graph representation by its subsets (subgraphs/communities) extends to general topological spaces. Using this framework, we rigorously define fractal networks and connect their properties with established combinatorial concepts, such as graph colorings and descriptive complexity.

The theoretical framework surveyed here sets a foundation for applications to real-life networks and future studies of fractal characteristics of complex networks using combinatorial methods and algorithms.

The goal of this paper is to introduce to the broader research community a theoretical and computational framework designed for robust and rigorous analysis of self-similarity and fractality in complex networks. The traditional definition of a fractal is an object whose Hausdorff (fractal) dimension is greater than its Lebesgue (topological) dimension. However, applying these concepts precisely to networks, particularly finite real-world networks, is challenging. This is because topology traditionally concerns continuous objects, which networks are not, and the concepts of fractality and self-similarity typically apply to infinite structures, unlike finite networks.

The theory we present here bridges these concepts with well-established, clear, and comprehensible combinatorial characteristics applicable to finite networks. For example, a network's self-similarity arises from the pattern of overlaps between its densely connected communities. The framework we have developed enables the rigorous definition, precise interpretation, and accurate estimation of various properties of fractal networks.

I. INTRODUCTION

To David- soul of chaos community

Over the past two decades, research focusing on self-similarity and fractality in discrete structures, such as particularly complex networks, has gained momentum[1–4]. These studies have yielded many valuable insights into the structure and development of complex networks. However, they frequently lack a precise and thorough mathematical basis, leading to uncertainties about what truly defines the self-similarity and fractality of networks. This gap often makes it challenging to interpret the results of such research effectively.

Various ways of adapting the concepts of self-similarity and fractality from general topology to discrete structures like complex networks are possible. According to Mandelbrot's classical definition, a geometric fractal is a topological space with the topological (Lebesgue) dimension being different from the fractal (Hausdorff) dimension. It is also commonly assumed that fractals exhibit a form of geometric self-similarity[5]. Therefore, when studying self-similarity and fractality of networks, it is crucial to develop appropriate discrete analogues of aforementioned concepts.

A common approach is to consider a network as a finite metric space, in which the distance between any two vertices is determined by the length of the shortest path connecting them. The network fractal dimension is then identified with the Minkowski–Bouligand (box-counting) dimension[1,2]. More specifically, for a network $\mathscr{G}$, suppose that $V(\mathscr{G}) = B_1 \cup \cdots \cup B_{N_l}$ is the minimal cover of $\mathscr{G}$ by l-boxes (subnetworks of diameter at most l). Then the *fractal dimension* $d = d(\mathscr{G})$ is defined by the relation

$$N_l \approx C \cdot l^{-d}, \tag{1}$$

where C is a constant. This definition is often applied under the assumption that it holds true for sufficiently large l.

While this approach is intuitive and easy to compute, it encounters several challenges in terms of mathematical rigor, in-

terpretation and applications. These challenges include:

1) The approximate nature of the definition (1), which is inherently asymptotic. This means that, strictly speaking, it is applicable to sequences of networks $\mathscr{G}_1,...,\mathscr{G}_n$ or random network generation process when the network sizes approach infinity, rather than to individual finite networks. However, in many cases, it is necessary to analyze the self-similarity and fractality of distinct finite networks that cannot be described by an asymptotic model. In such cases, reliable estimation of box-counting dimensions is problematic, because intrinsic finiteness and discretness of real networks prevents the accumulation of a sufficient number of data points to get reliable finite approximations of continuous functions. This problem has been previously noted in the literature[6,7].

2) Limitation in applications of the definition only to particular network models, with many important classes of networks failing to exhibit a well-defined fractal dimension. This is particularly true for random scale-free networks, where the number of boxes N_l decreases exponentially with l, thereby making the box-counting definition inapplicable. The common workaround of classifying these networks as non-self-similar seems more of a semantic solution than an exact rigorous one. Furthermore, even when a network's box-counting dimension can be determined, correlating it with the network's structural or topological characteristics remains a challenge.

3) Limited understanding of self-similarity. The box-counting definition aligns with the concept of self-similarity as preservation of network properties under a length-scale transformation[1]. Thus, it mostly describes *statistical* self-similarity. However, geometric fractals have a stronger characteristic in the form of *geometric* self-similarity: they consist of parts topologically similar to the entire structure, not just exhibiting similar features at different scales.

4) Lack of a matching definition of topological dimension, which is necessary for definition of fractal networks. The conventional topological network dimension is 1, while the box-counting dimension is usually greater than 1. This makes almost all graphs fractal in terms of Mandelbrot's definition. Such understanding of fractality is not practically useful.

The observations above indicate that straightforwardly applying continuous concepts to discrete objects, like networks, can pose challenges. Consequently, there is a significant need to cultivate an understanding of network dimensionality, self-similarity, and fractality rooted in the fundamental principles and methodologies inherent in graph theory and combinatorics.

In this paper, we overview such an approach, that was first introduced and developed in[8]. It establishes the direct graph-theoretical analogues of topological (Lebesgue) and fractal (Hausdorff) dimensions in a way that naturally links them to combinatorial parameters well known to graph theorists and first introduced to link graph theory with category theory. This approach allows to demonstrate that, roughly speaking, self-similarity of networks is defined by the patterns of intersection between dense network communities. Using this natural and intrinsic to networks framework, one can rigorously define fractal networks and link their properties to established combinatorial concepts, such as graph colorings and descriptive complexity.

II. TOPOLOGICAL AND FRACTAL DIMENSIONS OF NETWORKS

This section describes definitions and properties of graph-theoretical equivalents of Lebesgue and Hausdorff dimensions of topological spaces.

A. Basic definitions from graph theory.

In mathematical terms, a network is a graph. The following standard graph-theoretical definitions will be used throughout this paper. Consider a graph G with the vertex set $V(G)$ and the edge set $E(G)$. A graph H is a *subgraph* of G, if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. A subgraph $G[U]$ is *induced* by a vertex subset $U \subseteq V(G)$, if it contains all edges with both endpoints in U.

The notation $x \sim y$ is used to denote that the vertices x and y are adjacent (connected by an edge). A vertex degree $\deg(v)$ is a number of vertices adjacent to the vertex v. A vertex of degree 1 is called a *pendant vertex*. The maximum vertex degree of G is denoted by $\Delta(G)$.

A graph is *connected*, if there is a path between any pair of vertices; otherwise it is *disconnected*. In the latter case, a graph is a disjoint union of connected subgraphs called *connected components*.

We use $\overline{G}$ to denote the complement of G, i.e. a graph on the same vertex set where two vertices are adjacent if and only if they are not adjacent in G. The connected components of $\overline{G}$ are referred to as *co-connected components* of G. A graph is *2-connected* if it remains connected after the removal of any vertex. All graphs considered in this paper are assumed to be connected.

A complete graph is a graph where every pair of vertices is linked by an edge. The notations K_n, P_n, and C_n refer to the complete graph, the path and the cycle on n vertices, respectively. A *star* graph, noted as $K_{1,n}$, is a graph consisting of $n+1$ vertices where one vertex has a degree of n and the remaining n vertices are pendant.

A *clique* of G is a subset of its vertices that form a complete subgraph. A family of cliques $\mathscr{C} = (C_1,...,C_m)$ of G is a *clique cover*, if every edge $uv \in E(G)$ is contained in some clique from $\mathscr{C}$.

A hypergraph $\mathscr{H} = (\mathscr{V}(\mathscr{H}), \mathscr{E}(\mathscr{H}))$ extends the concept of a graph. In a hypergraph, the edge set $\mathscr{E}(\mathscr{H})$ can contain any subset of the vertex set $\mathscr{V}(\mathscr{H})$, i.e, an edge can join any number of vertices. In other words, the edge set of a hypergraph is a *set system*, i.e. a collection of subsets of a finite set $\mathscr{V}(\mathscr{H})$ The *rank* of a hypergraph, $r(\mathscr{H})$, is defined as the maximal size of its edges.

A hypergraph $\mathscr{H}$ can be transformed into a simple graph through the concept of an *intersection graph* $L = L(\mathscr{H})$. In this graph, each vertex v_E represents an edge E in $\mathscr{H}$, and

two vertices $v_E, v_F \in V(L)$ are adjacent if their corresponding edges intersect, i.e., $E \cap F \neq \emptyset$. If $G = L(\mathscr{H})$, then G is said to be *represented* by the set system $\mathscr{E}(\mathscr{H})$. In essence, G encodes the intersection pattern of the set system $\mathscr{E}(\mathscr{H})$, and, conversely, the vertices of G are encoded by the sets from $\mathscr{E}(\mathscr{H})$.

A hypergraph $\mathscr{H}$ is termed *strongly k-colorable* if its vertices can be colored using colors from the set $1,\ldots,k$ such that no two vertices within the same edge share the same color. The sets of vertices sharing a common color are referred to as *color classes*.

Other definitions will be introduced as needed.

B. Lebesgue dimension of networks

We begin by discussing the Lebesgue dimension, starting with its general topology definition. Let X be a compact metric space. A family $\mathscr{C} = \{C_\alpha : \alpha \in A\}$ of open subsets of X is termed a *cover* if $X = \bigcup_{\alpha\in A} C_\alpha$. A cover $\mathscr{C}$ is classified as a *k-cover* if each point $x \in X$ is contained in no more than k sets from $\mathscr{C}$. It is an *ε-cover* if the diameter of every set C_i in $\mathscr{C}$ is at most ε. If a cover is both an ε-cover and a k-cover, it is referred to as an (ε,k)-*cover*.

The *Lebesgue dimension* of a X, denoted as $dim_L(X)$, is defined as the minimal integer k such that for any $\varepsilon > 0$ there exists an $(\varepsilon, k+1)$-cover of X. So, the Lebesgue dimension of a metric space is determined using k-covers comprised of sets with arbitrarily small diameters. This concept can be naturally adapted to graphs by considering graph k-covers by subgraphs of smallest possible diameter, i.e. by cliques. The corresponding parameter is known in graph theory as a *rank dimension*[9] $dim_R(G)$, which is defined precisely as the minimal integer k such that G has a clique k-cover. Consequently, the Lebesgue dimension $dim_L(G)$ of a graph is equated to its rank dimension $dim_R(G)$ minus 1.

Interestingly, this association connects the Lebesgue dimension of a graph with graph representation by a particular type of set systems. This connection enables the establishment of relationships between the fractality of a graph and its descriptive complexity, as discussed in Section II C 4. Specifically, the following fact is a reformulation of a theorem due to C. Berge:

Theorem 1. [10] *$dim_L(G) \leq k$ if and only if G is an intersection graph of a hypergraph of rank $\leq k+1$.*

No less interestingly, this connection with set system representation can be extended from graph theory back to general topology, thereby reinforcing the validity of the proposed association. It has been demonstrated in[8] that compact metric spaces X share a similar property: $dim_L(X) \leq k-1$ if and only if X can be *approximately* represented, with any chosen level of accuracy, by a hypergraph of rank at most k in such a way that any two points of X are close to each other whenever the hypergraph edges corresponding to these points have a nonempty intersection. Formally, this property is established by the following theorem:

Theorem 2. [8] *Let X be a compact metric space with a metric ρ. Then $dim_L(X) \leq k$ if and only if for any $\varepsilon > 0$ there exists a number $0 < \delta < \varepsilon$ and a hypergraph $\mathscr{H}(\varepsilon)$ on a finite vertex set $V(\mathscr{H}(\varepsilon))$ with an edge set $E(\mathscr{H}(\varepsilon)) = \{e_x : x \in X\}$, which satisfies the following conditions:*

1) $rank(\mathscr{H}(\varepsilon)) \leq k+1$;

2) $e_x \cap e_y \neq \emptyset$ for every $x, y \in X$ such that $\rho(x,y) < \delta$;

3) $\rho(x,y) < \varepsilon$ for every $x, y \in X$ such that $e_x \cap e_y \neq \emptyset$;

4) for every $v \in V(\mathscr{H}(\varepsilon))$ the set $X_v = \{x \in X : v \in e_x\}$ is open.

C. Hausdorff dimension of networks.

Identifying the Hausdorff dimension of graphs presents a greater challenge, as the underlying continuous definition is more intricate. Specifically, the continuous Hausdorff dimension relies on the concept of a *measure*. Informally, a measure is a function that allocates a non-negative numerical "size" to particular subsets of a given set, extending traditional notions of length, area and volume. The measure (as any probability distribution) should be countably additive, i.e. the measure of a countable union of disjoint sets equals the sum of their individual measures.

The *d-dimensional Jordan measure* defines the volume of a bounded set X in Euclidean space as the limit of the volumes of finite covers of X by disjoint hyperrectangles, where a d-dimensional hyperrectangle $R = [a_1, b_1) \times \cdots \times [a_d, b_d)$ is a Cartesian product of semi-open intervals. The *d-dimensional Lebesgue measure* is defined similarly, with the limit taken over all countable covers of X by (not necessarily disjoint) hyperrectangles. Finally, the *d-dimensional Hausdorff measure* of the set X is given as $\mathscr{H}^d(X) = \lim_{\varepsilon\to 0} \mathscr{H}^s_\varepsilon(X)$, where $\mathscr{H}^d_\varepsilon(X) = \inf\{\sum_{C\in\mathscr{C}} diam(C)^d\}$, and the infimum is taken over all ε-covers of X. These 3 measures are related: the Jordan and Lebesgue measures of set X are equal, if the former exists, while Lebesgue and Hausdorff measures of so-called Borel sets (i.e. sets formed by intersections, complements and countable unions of open sets) differ only by a constant factor.

Hausdorff dimension $dim_H(X)$ of the set X is defined as the threshold at which the Hausdorff measure of X transitions from infinity to a finite value:

$$dim_H(X) = \inf\{s \geq 0 : \mathscr{H}^s(X) < \infty\}. \tag{2}$$

The Hausdorff dimension is particularly effective in calculating the measures of self-similar sets[5]. Additionally, it has a connection to the descriptive complexity of various mathematical objects[11,12]. Finally, Lebesgue and Hausdorff dimension of are related as follows:

$$dim_L(X) \leq dim_H(X). \tag{3}$$

Considering these definitions and properties, a well-founded graph analogue of the continuous Hausdorff dimension should satisfy the following criteria:

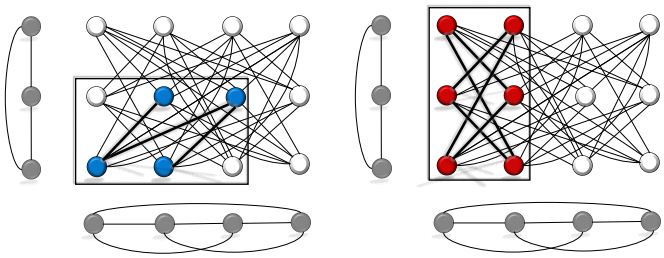

FIG. 1. The embeddings of the graphs P_4 (a path with 4 vertices) and C_6 (a cycle with 6 vertices) into the 2-dimensional space formed by the tensor product $K_4 \times K_3$, together with their respective minimal covers by hyperrectangles, each having a volume of 6. Consequently, the volumes of both graphs P_4 and C_6 are determined to be 6.

1) It should be related to the graph Lebesgue dimension as in (3).

2) It should be defined in relation to an equivalent of the Hausdorff measure for graphs.

3) It should be associated with self-similarity of graphs.

4) It should be linked to the descriptive complexity of graphs.

It turned out that the graph parameter satisfying all these requirements do exist: it is so-called *product dimension*[13–15], also referenced in various sources as *Prague dimension*[16] or *Nešetřil-Rödl* dimension[17]. Originally introduced in the 1970s, it emerged in the context of attempts to describe graphs in terms of abstract algebra.

Tensor product of graphs G_1 and G_2 is the graph $G_1 \times G_2$ with the vertex set $V(G_1 \times G_2) = V(G_1) \times V(G_2)$ with two vertices (u_1, u_2) and (v_1, v_2) being adjacent whenever u_1 and v_1 are adjacent in G_1, and u_2 and v_2 are adjacent in G_2. *Product dimension* $dim_P(G)$ is the minimal integer d such that G is an induced subgraph of a tensor product of d complete graphs[13].

It is posited that the Hausdorff dimension of a graph can be equated to the Prague dimension of its complement, minus 1: $dim_H(G) = dim_P(\overline{G}) - 1$. In what follows, we will substantiate this association.

Several equivalent definitions of product dimension, and consequently, of Hausdorff dimension, exist[13,18–21]. In our context, it is important that the Hausdorff dimension, just as Lebesgue dimension, can be associated with clique covers, as well as with graph representation by set systems (see Theorem 3).

An *equivalent k-cover* of a graph G is a cover composed of k subgraphs, each being a disjoint union of cliques. This can also be defined as a clique cover where the cliques are colored in k colors so that intersecting cliques have different colors. A clique cover is *separating*, if for any two distinct vertices there exists at least one clique that includes one vertex but not the other.

The following theorem summarizes slightly reformulated results obtained in previous studies:

Theorem 3. [13,18,20] *The following statements are equivalent:*
a) $dim_H(G) \leq k$.
b) There exists a separating equivalent $k+1$*-cover of* G*;*
c) G *is an intersection graph of strongly* $k+1$*-colorable hypergraph without multiple edges;*
d) G *can be embedded into the* $k+1$*-dimensional integer grid in such a way that two vertices are adjacent if and only if the corresponding grid points share a coordinate.*

Using Theorem 3, it can be shown that our definition of graph Hausdorff dimension indeed meets the conditions 1) - 4).

1. Relation between Lebesgue and Hausdorff dimensions of graphs allows to identify fractal graphs.

This relation immediately follows from Theorem 1 and Theorem 3c, since by definition, any strongly $(k+1)$-colorable hypergraph has a rank of at most $k+1$. This allows to define fractal graphs in a manner analogous to that for topological spaces: a graph G is a *fractal* if

$$dim_L(G) < dim_H(G). \tag{4}$$

For instance, the order 2 and order 3 Sierpinski gasket graphs S_2 and S_3 depicted on Fig. 2 and Fig. 3 satisfy this definition: each vertex of these graphs is covered by two cliques, indicating that $dim_L(S_2) = dim_L(S_3) = 1$, while the cliques can be colored using three colors, showing that $dim_H(S_2) = dim_H(S_3) = 2$. As a matter of fact, it can be shown that a Sierpinski gasket graph of any order is a fractal graph , with Lebesgue and Hausdorff dimensions equal to 1 and 2, respectively[8]. Fractal graphs are discussed in more detail in Section III.

2. Hausdorff dimension is associated with a network measure.

The key problems in this context are twofold: firstly, identifying an appropriate analogue for the "volume" of a network, and secondly, determining if this "volume" really defines the Hausdorff dimension as we understand it. To address these questions, we introduce and explore a network equivalent of the Jordan measure (which, in a finite topology, is equivalent to the Hausdorff measure) and then establish its connection to the Hausdorff dimension.

In our approach, the graph equivalent of Euclidean space is conceptualized as the tensor product of d complete graphs $(K_n)^d = K_n \times \cdots \times K_n$. For simplicity, we can assume that the vertices of each complete graph are the integers from 1 to n. The *d-dimensional rectangle* R is a subgraph of $(K_n)^d$, that is defined as a tensor product of complete subgraphs $R = K_n[J_1] \times \cdots \times K_n[J_d]$ (Fig. 1), where $J_1, ..., J_d$ are non-empty

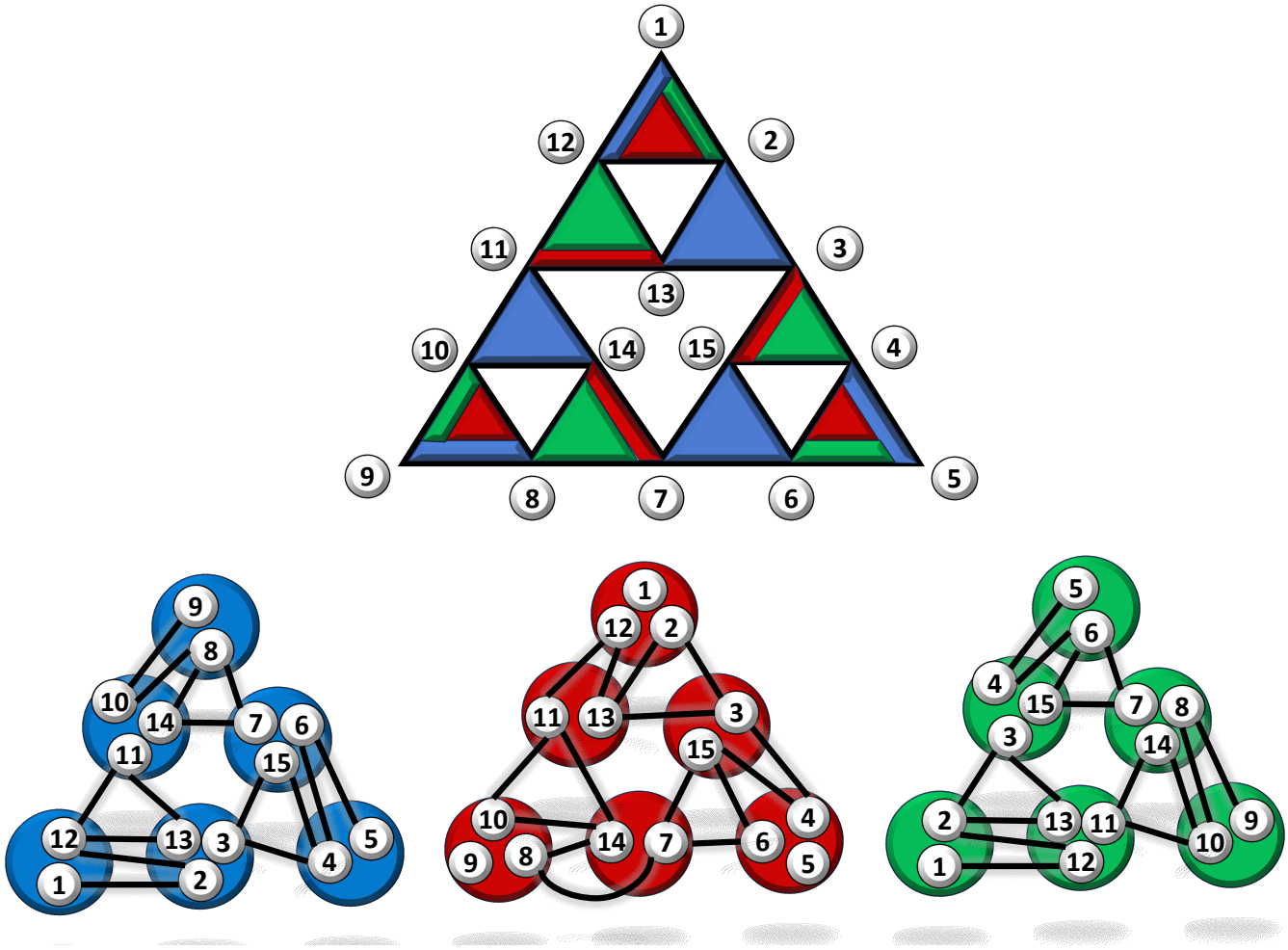

FIG. 2. Self-similarity of S_3, the Sierpinski gasket graph of order 3. **Top:** an equivalent separating 3-cover. The cliques comprising the cover are color-coded in red, blue, and green; cliques of the same color are non-intersecting, as required by the definition. Consequently, the Hausdorff dimension of S_3 is determined to be 2. **Bottom:** contractions of S_3 produced by similarity mappings associated with each color. Each contraction results from merging the vertices of cliques of the given color; in addition, we depict non-contracted edges. These contractions are each isomorphic to a Sierpinski gasket graph of order 2, and the union of non-contracted edges from all three contractions produces the original graph S_3, thus exemplifying its self-similar nature.

subsets of $\{1,...,n\}$. The *volume* of R is naturally determined as $vol(R) = \prod_{i=1}^{d} |J_i|$ (Fig. 1).

It is known, that every graph G is can be embedded into $(K_n)^d$ with sufficiently large d and n[13]. Given such an embedding, a set of disjoint rectangles $\mathscr{R} = \{R^1,...,R^m\}$ constitutes a *rectangle co-cover* of G, if it covers all vertices and all pairs of non-adjacent vertices of G (Fig. 1). The *d-measure* of a graph G is defined as follows:

$$\mathscr{H}^d(G) = \begin{cases} \min\limits_{\overline{G}'} \min\limits_{\mathscr{R}} \sum\limits_{R\in\mathscr{R}} vol(R), \overline{G} \text{ is embeddable into } (K_n)^d; \\ +\infty, \text{ otherwise} \end{cases} \tag{5}$$

Here, the first minimum is taken over all embeddings $\overline{G}'$ of the complement graph $\overline{G}$ into $(K_n)^d$, and the second minimum – over all rectangle co-covers of an embedding. For instance, Fig. 1 shows that the 2-dimensional volumes of the path P_4 and the cycle C_6 are both equal 6. The key task now is to prove that this defined function indeed qualifies as a measure, i.e. it is additive. This is established by the following theorem:

Theorem 4. [8] *Let G_1 and G_2 be two graphs, and $G_1 \cup G_2$ be their disjoint union. Then*

$$\mathscr{H}^d(G_1 \cup G_2) = \mathscr{H}^d(G_1) + \mathscr{H}^d(G_2) \tag{6}$$

Thus, similarly to the continuous definition, we can define a Hausdorff dimension of a graph G as the minimal integer d for which $\mathscr{H}^d(G)$ is finite. This leads precisely to the product dimension of $\overline{G}$.

In addition, Hausdorff measure can be linked to network descriptive complexity (see Subsection II C 4).

3. Hausdorff dimension is associated with self-similarity of networks.

Informally, a metric space X with a metric d is self-similar, if it is comprised of parts similar to itself. Formally, the self-similarity can be defined using the notion of a *contraction*[22]. A mapping $f : X \to X$ is a contraction, if $d(f(u), f(v)) \leq \alpha d(u,v)$ for all $u, v \in X$, where $\alpha < 1$ is called its *similarity ratio*. The space X is *self-similar*, if there exists a family of contractions $f_1, ..., f_k$ such that $X = \bigcup_{i=1}^{k} f_i(X)$.

This continuous definition cannot be directly applied to discrete metric spaces such as networks, since for them contractions in the aforementioned sense do not exist. Instead, the self-similarity of networks can be rigorously defined as follows. In this section, it is convenient to assume that every vertex is adjacent to itself. For two graphs G and H, a *homomorphism*[13] is a mapping $f : V(G) \to V(H)$ which maps adjacent vertices to adjacent vertices. A homomorphism f is a *similarity mapping*, if inverse images of adjacent vertices are also adjacent. In other words, similarity mapping contracts some edges in such a way that images and inverse images of cliques remain cliques. With a similarity mapping f we can associate a subgraph G^f of G, which is formed by all edges that are not contracted (Fig. 2).

A family of graph similarity mappings $f_1, ..., f_k$ is a *contracting family*, if every edge of G is contracted by some mapping. Then a graph G is *self-similar*, if $G = \bigcup_{i=1}^{k} G^{f_i}$ (Fig. 2). It turned out that network self-similarity defined in such a way is directly associated with its Hausdorff dimension:

Theorem 5. [8] *Graph G is self-similar with a contracting family of k similarity mappings if and only if $dim_H(G) \leq k$. Furthermore, these mappings can be constructed from the corresponding equivalent separating k-cover, whose existence is guaranteed by Theorem 3.*

The concept of self-similarity applied to the Sierpinski gasket graph S_3 is illustrated by Fig. 2.

In general, a family of similarity mappings satisfying Theorem 5 is not unique; therefore, it is logical to search for the most parsimonious similarity mappings. The degree of a graph's self-similarity can be quantified by the smallest number of similarity mappings in a contracting family, i.e. by its Hausdorff dimension. A smaller number of similarity mappings signifies a more compact packing of a network by its contractions, indicating a higher degree of self-similarity. For the same reasons, when there are multiple families of similarity mappings of minimal size, it makes sense to focus on those with the smallest images, that is, those with the fewest number of cliques associated with each color in the corresponding equivalent separating k-covers. For comparing the self-similarity of networks of varying sizes, the Hausdorff dimension can be normalized.

4. Lebesgue and Hausdorff dimensions are linked to the descriptive complexity of networks.

Theorems 1 and 3 provide a framework to understand graph Lebesgue and Hausdorff dimensions, as well as fractality, from the information theory perspective. Consider a graph G with Lebesgue dimension $dim_L(G) = k$. According to this, G is represented by a set system of rank $k+1$. Therefore, each vertex in G is represented by $k+1$ "coordinates," comprising the corresponding set in the system. Notably, these coordinates are unordered, and the adjacency of a pair of vertices u and v is defined by the existence of a common coordinate between u and v (see Fig. 3).

Similarly, Theorem 3d posits that graphs with a Hausdorff dimension $dim_H(G) = k$ are characterized by $k+1$ coordinates as well. However, in this case, the coordinates are ordered *vectors*, and the adjacency of two vertices is identified by the presence of a shared coordinate in the same position.

In this context, non-fractal networks are graphs where set and vector representations are equivalent, meaning that the additional information provided by the order of coordinates is not necessary for their description. In contrast, fractal networks possess extra structural features that necessitate the use of additional information contained in vector representations for their complete encoding. This concept is visually exemplified in Fig. 3.

The connection between network Hausdorff dimension and information complexity can be further explored using the concept of *Kolmogorov complexity*. In simple terms, the Kolmogorov complexity $K(s)$ of a string s is the length of the shortest encoding of that string that allows to reconstruct it without any loss of information[23].

Every network can be encoded using the string representation of its adjacency matrix. The Kolmogorov complexity $K(G)$ of a network G can thus be defined as the Kolmogorov complexity of this string[24]. Note that a network G and its complement $\overline{G}$ contain the same information; therefore, an encoding for one encodes the other as well.

The definition directly implies that $K(G) = O(n^2)$. An n-vertex network can be also represented by a list of its edges, with the endpoints of each edge encoded in binary form and concatenated with the binary representation of n. This approach leads to the estimation $K(G) \leq 2m\log(n) + \log(n) = O(m\log(n))$[23,24].

The Hausdorff measure and dimension of a graph G provide means to derive more detailed estimates of its Kolmogorov complexity. Suppose that G has a Hausdorff dimension d, and both G and its complement $\overline{G}$ are connected (otherwise, each connected or co-connected component can be encoded separately). In this case, G has a finite measure $\mathscr{H}^d(G)$ determined by the volume of $\overline{G}$ embedded in the tensor product of d complete graphs. It can be shown[8] that the minimal rectangle co-cover of a graph that is both connected and co-connected consists of a single rectangle

$$K_{p_1} \times \cdots \times K_{p_d}, \tag{7}$$

where $\mathscr{H}^d(G) = p_1 \cdot \ldots \cdot p_d$.

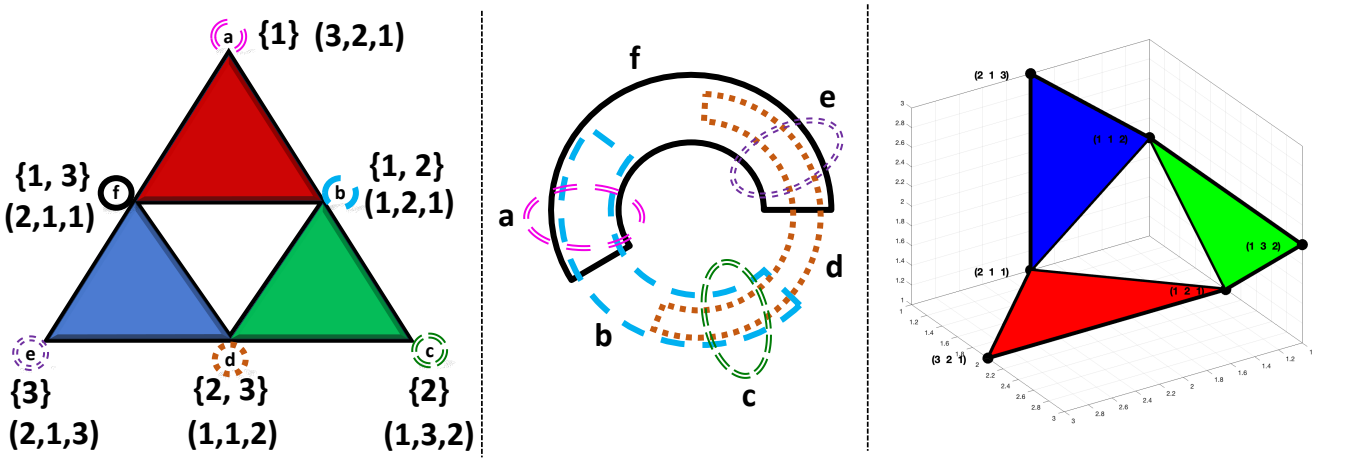

FIG. 3. **Left Panel:** the Sierpinski gasket graph of order 2, S_2, with a cover by 3 cliques that are colored in red, blue, and green. Each vertex of S_2 is covered by two cliques, indicating that $dim_L(S_2) = 1$. The cliques are colored using three colors, suggesting that $dim_H(S_2) = 2$. This combination of Lebesgue and Hausdorff dimensions shows that S_2 is a fractal. **Middle Panel:** a schematic illustration of how S_2 is represented by a set system of rank 2. Alongside each vertex in the left panel, the corresponding set that encodes it is displayed. **Right Panel:** embedding of S_2 into a 3-dimensional grid. The vector representing each vertex is indicated next to the corresponding vertex in the left panel.

Thus, using the embedding of $\overline{G}$ into (7), $\overline{G}$ (and consequently G) can be encoded by assigning each vertex v a vector of its coordinates in this embedding $\phi(v) = (\phi_1(v), \ldots, \phi_d(v))$ with $\phi_j(v) \in 1, \ldots, p_j$. This encoding can be represented as a string containing the binary representations of the coordinates $\phi_j(v)$, each using $\log(p_j)$ bits, and concatenated with the binary representations of integers n and $p_1, \ldots, p_d$. The length of this string is:

$$(n+1)\sum_{j=1}^{d}\log(p_j)+\log(n) = (n+1)\log(\mathscr{H}^d(G))+\log(n). \tag{8}$$

This leads to the following estimate:

Proposition 1.

$$K(G) \leq (n+1)\log(\mathscr{H}^d(G))+\log(n) \tag{9}$$

A rougher estimate can be derived by noting that $p_j \leq n$ due to the minimality of the rectangle (7). Consequently,

$$K(G) \leq (n+1)d\log(n) = O(dn\log(n)). \tag{10}$$

thereby indicating that the Hausdorff dimension can serve as a measure of a network's descriptive complexity.

III. FRACTAL NETWORKS

Recall that fractal networks are the networks for which their Hausdorff dimension differ from Lebesgue dimension. A specific example of such fractal networks, the Sierpinski gasket graphs, was examined in previous sections. In this section, we discuss other types of fractal networks and specifically illustrate how network fractality connects to the classic graph-theoretical concept of edge coloring.

A. Fractality as a generalization of edge coloring dichotomy.

The concept of *edge k-coloring* is similar to vertex coloring, requiring that edges sharing a common vertex receive different colors. The *edge chromatic number* of a graph G, denoted as $\chi'(G)$, is the smallest number of colors needed for its edge coloring. A classical theorem by V. Vizing[25] establishes that the edge chromatic number of any graph can only be either $\Delta(G)$ or $\Delta(G)+1$, where $\Delta(G)$ is the graph's maximum degree. This leads to a famous dichotomy that classifies all graphs into two categories: class 1, where $\chi'(G) = \Delta(G)$, and class 2, where $\chi'(G) = \Delta(G)+1$. It is worth noting that despite this fact, determining the exact edge chromatic number of a given graph is a computationally hard algorithmic problem[17]; furthermore, a similar dichotomy does not exist for vertex coloring[21].

Separation of networks into non-fractals and fractals is a generalization of the class 1/class 2 dichotomy. Essentially, edge k-coloring can be viewed as a specific instance of a separating equivalent k-cover, where each clique is an edge. In the case of triangle-free graphs (those without cliques of size 3), this type of separating equivalent k-cover is the only possible option. Therefore, triangle-free fractal networks are precisely triangle-free graphs of class 2.

However, the connection between fractality and edge coloring goes deeper than that. A notable link can be established for cubic graphs, i.e. graphs with all vertices having the degree of three. Cubic graphs are not only studied in graph theory, but also arise in topology, physics and analysis of complex networks[26,27].

A notable category of cubic fractals is exemplified by the remarcable class of *snarks*[28,29]. A snark is defined as a 2-connected, triangle-free cubic graph of class 2. Snarks represent a significant and long-studied class of graphs, whose structural properties are still incompletely understood[29]. The

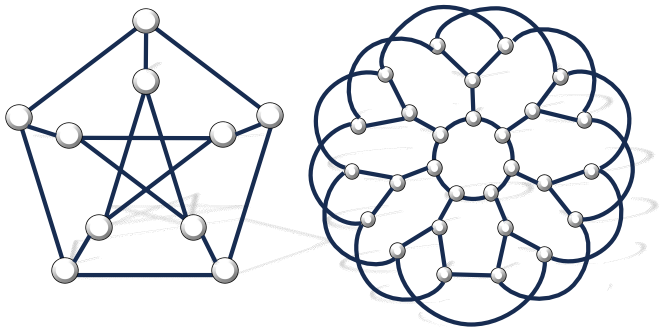

FIG. 4. **Examples of snarks.** Left: Petersen graph. Right: flower snark[31].

discovery of a new non-trivial snark is considered a substantial scientific achievement (hence the term coined by M. Gardner[28]), paralleling the discovery of new fractals. Many known snarks exhibit a high degree of symmetry and are constructed through recursive procedures, as shown in Fig. 4. This analogy is quite apt since snarks are indeed fractal graphs with a Hausdorff dimension of $\dim_H(G) = 2$. Moreover, it can be demonstrated that snarks are the foundational cubic fractals, in that every 2-connected cubic fractal can be transformed into a snark through a sequence of elementary operations, including the removal of triangles and the subsequent contraction of pendant vertices created by this process[8].

It should be also noted that the relation between graph Lebesgue and edge chromatic number implies that calculating the former is an NP-hard (i.e. algorithmically hard) problem. It follows from the NP-hardness of the edge chromatic number problem for triangle-free cubic graphs[30].

B. Fractality of random networks.

The concept of network fractality extends the dichotomy of edge coloring beyond cubic graphs to encompass standard random network models, including scale-free graphs and Erdős-Rényi graphs. For these types of networks, it is possible to establish analogues of Vizing theorem.

A *scale-free network* is characterized by a degree distribution that asymptotically follows a power law, meaning the probability that a vertex has a degree d can be approximated by the function $ad^{-\alpha}$, where a is a constant and α is the *scaling exponent*. Scale-free networks are among the most extensively studied types of complex networks, with numerous examples found in physics, social sciences, biology, medicine and other fields[32]. Most models of scale-free networks are based on variants of the preferential attachment scheme, where new vertices preferentially connect to existing higher-degree vertices. An alternative mathematically rigorous model proposed in[33,34] describes an n-vertex scale-free network $G(n,\alpha)$ with scaling exponent α, constructed by assigning an ith vertex a weight $w_i = (\frac{n}{i})^{1/\alpha}$ and connecting each pair of vertices i, j by an edge with the probability $p_{ij} = 1 - e^{-\frac{w_i w_j}{n}}$.

For scale-free graphs built according to this model, there is an analogue of Vizing theorem. To formulate it, we define a random event as occurring with high probability if its probability approaches 1 as $n \to \infty$.

Theorem 6. [8] *For a scale-free graph $G = G(n,\alpha)$, where $\alpha > \frac{12}{5}$, with high probability the Lebesgue dimension $\dim_L(G)$ is either $\Delta(G)-2$ or $\Delta(G)-1$, while the Hausdorff dimension $\dim_H(G)$ can take one of four possible values: $\Delta(G)-2, \Delta(G)-1, \Delta(G), \Delta(G)+1$.*

Erdős-Rényi random graphs, denoted as $G(n,p)$, are generated by independently adding edges between pairs of vertices with a probability p. It has been demonstrated that the conclusions of Theorem 6 are applicable to sparse Erdős-Rényi random graphs where $p = \frac{1}{n^\alpha}$ and $\alpha > \frac{5}{6}$[8].

Combining Theorem 6 with known asymptotic of maximal degrees[35–37] provides insights into the asymptotic behavior of the Hausdorff dimension of scale-free graphs produced by the preferential attachment scheme and of sparse Erdős-Rényi graphs. Specifically, with high probability $\dim_H(G) = O(\sqrt{n})$ for preferential attachment graphs and $\dim_H(G) = O(\log n)$ for sparse Erdős-Rényi graphs $G = G(n, \frac{1}{n^\alpha})$ with appropriate α. For dense Erdős-Rényi graphs, the asymptotic behavior of their Hausdorff dimension has been described in[16]:

Theorem 7. [16] *For any fixed edge probability p, there are constants c and C such that with high probability*

$$c\frac{n}{\log n} - 1 \le \dim_H(G(n,p)) \le C\frac{n}{\log n} - 1. \qquad (11)$$

IV. EXPERIMENTAL APPLICATIONS

When applying the described framework to real-world complex networks, it is crucial to acknowledge that the data used to construct these networks often contains noise. Consequently, the likelihood of identifying meaningful communities as perfect cliques in real networks is low. Therefore, in practical computations, it is advisable to relax the model's constraints by substituting cliques with dense network communities identified using one of many community detection algorithms[38]. Furthermore, focusing on dense communities enhances the robustness of estimations, since a clique can be disrupted by the absence of a single edge, while a dense community exhibits greater resilience to edge removal.

Utilizing this approach, we calculated the Lebesgue and Hausdorff dimensions of several real-life networks[8]. A particularly notable result was that approximately one third of these networks were identified as fractals. While our sample size was limited, it is compelling to observe that such prevalence of fractals notably exceeds the predictions made by theoretical estimations for the network models previously discussed. This finding implies that these models might not fully capture intricate topological features of real networks, such as self-similarity and fractality, indicating a potential need for more refined and nuanced modeling approaches.

An interesting example from the field of biology involves the analysis of phylogenetic networks used to represent the structure of biological populations. In such networks, vertices represent the genomes of population members, with adjacency indicating genetic closeness between genomes.

We analyzed the Hausdorff dimensions of genetic networks from Hepatitis C virus populations sampled from infected individuals during acute and chronic stages of infection[8]. Our findings revealed that the Hausdorff dimensions for chronic stage populations are statistically significantly lower compared to those in acute stages. This outcome has biological significance as it implies a gradual evolution of intrahost viral populations towards increased self-similarity over the course of the infection. This trend indicates a progressive self-organization within viral populations, leading to the emergence of structural patterns in their composition. This points to a dynamic mechanism driving their formation in later infection stages, possibly linked to a higher level of adaptation and specialization among viral variants.

This observation is consistent with previously reported reductions in the Kolmogorov complexity of viral populations (measured for genomes rather than for phylogenetic networks)[39]. It also corroborates recent models of viral antigenic cooperation[40–43], suggesting that viral populations can exhibit complex behaviors and adapt to host environments as quasi-social systems.

V. CONCLUDING REMARKS

The concepts initially introduced in[8] and discussed in this paper are expected to be useful for theoretical explorations of network models, as well as for analyzing experimental networks encountered in various research areas. Moreover, identifying and examining fractal properties in networks can now be approached as well-defined algorithmic problems. Examples of such problems include recognizing fractal networks and computing measures that indicate the extent to which a given network approaches fractality.

Another significant area for future research is gaining a more profound understanding of the structural characteristics of fractal networks and identifying network growth models that are capable of generating fractal structures. A type of model that shows promise in this area are the specialization models of network growth[44]. These models are based on the principle that as a network expands, it undergoes specialization in certain functions it is meant to perform. The model links the components involved in this specialization process to network motifs (statistically significant subnetworks) responsible for specific functions. These motifs are replicated across the network through a predefined specialization process, resulting in a network that exhibits a modular and self-similar structure. However, there are undoubtedly other approaches that can also be explored.

Finally, the proposed framework can be generalized by substituting cliques with dense subgraphs and by the extending it to directed graphs. These generalizations hold significant value for practical applications, yet the development of corresponding rigorous methodologies remains an ongoing task.

ACKNOWLEDGMENTS

The work of P.S. was partially supported by the NSF grants CCF-2047828 and IIS-2212508.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

[1] C. Song, S. Havlin, and H. A. Makse, "Self-similarity of complex networks," Nature **433**, 392–395 (2005).
[2] O. Shanker, "Defining dimension of a complex network," Modern Physics Letters B **21**, 321–326 (2007).
[3] S. N. Dorogovtsev and J. F. Mendes, *Evolution of networks: From biological nets to the Internet and WWW* (OUP Oxford, 2013).
[4] M. E. Newman, "The structure and function of complex networks," SIAM review **45**, 167–256 (2003).
[5] K. Falconer, *Fractal geometry: mathematical foundations and applications* (John Wiley & Sons, 2004).
[6] C. Song, L. K. Gallos, S. Havlin, and H. A. Makse, "How to calculate the fractal dimension of a complex network: the box covering algorithm," Journal of Statistical Mechanics: Theory and Experiment **2007**, P03006 (2007).
[7] Y. Xu, A. J. Gurfinkel, and P. A. Rikvold, "Architecture of the florida power grid as a complex network," Physica A: Statistical Mechanics and its Applications **401**, 130–140 (2014).
[8] P. Skums and L. Bunimovich, "Graph fractal dimension and the structure of fractal networks," Journal of Complex Networks **8**, cnaa037 (2020).
[9] Y. Metelsky and R. Tyshkevich, "Line graphs of helly hypergraphs," SIAM Journal on Discrete Mathematics **16**, 438–448 (2003).
[10] C. Berge, *Hypergraphs: combinatorics of finite sets*, Vol. 45 (Elsevier, 1984).
[11] B. Y. Ryabko, "The complexity and effectiveness of prediction algorithms," Journal of Complexity **10**, 281–295 (1994).
[12] L. Staiger, "Kolmogorov complexity and hausdorff dimension," Information and Computation **103**, 159–194 (1993).
[13] P. Hell and J. Nešetril, *Graphs and homomorphisms, volume 28 of Oxford Lecture Series in Mathematics and its Applications* (Oxford University Press, 2004).
[14] J. Nešetřil and V. Rödl, "A simple proof of the galvin-ramsey property of the class of all finite graphs and a dimension of a graph," Discrete Mathematics **23**, 49–55 (1978).
[15] L. Lovász, J. Nešetšil, and A. Pultr, "On a product dimension of graphs," Journal of Combinatorial Theory, Series B **29**, 47–67 (1980).
[16] H. Guo, K. Patton, and L. Warnke, "Prague dimension of random graphs," Combinatorica **43**, 853–884 (2023).
[17] M. Garey and D. Johnson, "Computers and intractability. a guide to the theory of np-completeness," (1979).
[18] N. Alon, "Covering graphs by the minimum number of equivalence relations," Combinatorica **6**, 201–206 (1986).
[19] N. Alon and R. Alweiss, "On the product dimension of clique factors," European Journal of Combinatorics **86**, 103097 (2020).
[20] A. Babaitsev and R. Tyshkevich, "K-dimensional graphs (in russian)," Vestnik Akademii nauk Belarusi , 75–81 (1996).
[21] D. West, *Introduction to graph theory* (Prentice Hall, 1996).
[22] G. Edgar, *Measure, topology, and fractal geometry* (Springer Science & Business Media, 2007).
[23] M. Li and P. Vitányi, *An introduction to Kolmogorov complexity and its applications* (Springer Science & Business Media, 2009).
[24] A. Mowshowitz and M. Dehmer, "Entropy and the complexity of graphs revisited," Entropy **14**, 559–570 (2012).
[25] V. G. Vizing, "On an estimate of the chromatic class of a p-graph," Discret Analiz **3**, 25–30 (1964).

[26]F. G. Woodhouse, A. Forrow, J. B. Fawcett, and J. Dunkel, "Stochastic cycle selection in active flow networks," Proceedings of the National Academy of Sciences **113**, 8200–8205 (2016).
[27]C. McDiarmid and F. Skerman, "Modularity of regular and treelike graphs," Journal of Complex Networks **6**, 596–619 (2017).
[28]M. Gardner, "Mathematical games," Scientific American **236**, 121–127 (1977).
[29]M. Chladnỳ and M. Škoviera, "Factorisation of snarks," the electronic journal of combinatorics **17**, 32 (2010).
[30]I. Holyer, "The np-completeness of edge-coloring," SIAM Journal on computing **10**, 718–720 (1981).
[31]R. Isaacs, "Infinite families of nontrivial trivalent graphs which are not tait colorable," The American Mathematical Monthly **82**, 221–239 (1975).
[32]M. Newman, *Networks* (Oxford university press, 2018).
[33]S. Janson, T. Łuczak, and I. Norros, "Large cliques in a power-law random graph," Journal of Applied Probability **47**, 1124–1135 (2010).
[34]F. Chung and L. Lu, "Connected components in random graphs with given expected degree sequences," Annals of combinatorics **6**, 125–145 (2002).
[35]A. Flaxman, A. Frieze, and T. Fenner, "High degree vertices and eigenvalues in the preferential attachment graph," Internet Mathematics **2**, 1–19 (2005).
[36]B. Bollobás and O. M. Riordan, "Mathematical results on scale-free random graphs," Handbook of graphs and networks: from the genome to the internet , 1–34 (2003).
[37]B. Bollobás, *Random graphs*, 73 (Cambridge university press, 2001).
[38]H. Cherifi, G. Palla, B. K. Szymanski, and X. Lu, "On community structure in complex networks: challenges and opportunities," Applied Network Science **4**, 1–35 (2019).
[39]P. B. Icer Baykal, J. Lara, Y. Khudyakov, A. Zelikovsky, and P. Skums, "Quantitative differences between intra-host hcv populations from persons with recently established and persistent infections," Virus Evolution **7**, veaa103 (2021).
[40]P. Skums, L. Bunimovich, and Y. Khudyakov, "Antigenic cooperation among intrahost hcv variants organized into a complex network of cross-immunoreactivity," Proceedings of the National Academy of Sciences **112**, 6653–6658 (2015).
[41]P. Domingo-Calap, E. Segredo-Otero, M. Durán-Moreno, and R. Sanjuán, "Social evolution of innate immunity evasion in a virus," Nature microbiology **4**, 1006 (2019).
[42]A. Leeks, L. M. Bono, E. A. Ampolini, L. S. Souza, T. Höfler, C. L. Mattson, A. E. Dye, and S. L. Díaz-Muñoz, "Open questions in the social lives of viruses," Journal of Evolutionary Biology **36**, 1551–1567 (2023).
[43]A. Leeks and S. West, "Altruism in a virus," Nature Microbiology **4**, 910–911 (2019).
[44]L. A. Bunimovich, D. Smith, and B. Z. Webb, "Specialization models of network growth," Journal of Complex Networks **7**, 375–392 (2019).