Complex Network Dimension and Path Counts

O. Shanker

Hewlett Packard Company, 16399 W Bernardo Dr., San Diego, CA 92130, U. S. A.

Theoretical Computer Science, To Be Published

Abstract

Large complex networks occur in many applications of computer science. The complex network zeta function and the graph surface function have been used to characterize these networks and to define a dimension for complex networks. In this work we present three new results related to the complex network dimension. First, we show the relationship of the concept to Kolmogorov complexity. Second, we show how the definition of complex network dimension can be made more rigorous by defining the concept for a single node, and then defining the complex network dimension as the supremum over all nodes. This makes the concept work better for formally infinite graphs. Third, we study interesting parallels to zeta dimension, a notion originally from number theory which has found connections to theoretical computer science. These parallels lead to a deeper insight into the complex network dimension, e.g., the formulation in terms of the entropy and a theorem relating dimension to connectivity.

Key words: Complex Networks, Graphs, Fractal Dimension

1 Introduction

Large complex networks occur in several diverse applications of computer science. Thus, measures relating to complex networks are of interest in computer science. The dimension of a complex network (in the large system limit) can be defined in different ways [14,15]. One common characteristic of dimension is that the higher the dimension of a system, the more complex the system

Email addresses: oshanker@gmail.com,
http://sites.google.com/site/complexnetworkdimension/ (O. Shanker).

is. For graphs, one measure of complexity is the dependence of the average number of paths on the length of the path. In section 2 we show the relation of the definition of complex network dimension to the Kolmogorov complexity of the graph [10] as defined in computational complexity theory. In section 3 we show how the definition of complex network dimension can be made more rigorous by defining dimension for a single node, and then defining the complex network dimension as the supremum over all nodes. In section 4 we relate the definition of the dimension to an expression similar to entropy rates [16,17] as used in computer science studies of dimension in discrete contexts. For graphs we show that the analogous relation is to the number of paths. In section 5 we show that a theorem for zeta dimension relating connectedness and dimension can be extended to complex network dimension with minor modification in the notation in the statement and the proof. We finally present the conclusions.

2 Relation to Kolmogorov Complexity

In this section we will show a relation between the complex network dimension and Kolmogorov complexity. The complex network zeta function [14] and the graph surface function [15] were introduced to characterize large graphs. They can be used to define the dimension of a complex network, if one starts with an arbitrarily large finite graph and defines the infinite limit by letting the size tend to ∞. We assume that the number of nearest neighbours for any node in the graph remains bounded as the graph size tends to ∞. If we define the average degree $\langle k \rangle$ of the graph as the average over all nodes of the number of nearest neighbours, then this means that the the average degree $\langle k \rangle$ is finite. As an example, consider the regular discrete d-dimensional lattice $\mathbf{Z}^d$ with L nodes in each dimension as a graph with the edges being the links between the nodes along the coordinate axes. The infinite size limit consists of letting L tend to ∞. The definition of complex network dimension should give the value d in this case.

In a graph, we denote by r_{ij} the distance from node i to node j, i.e., the length of the shortest path connecting the first node to the second node. r_{ij} is ∞ if there is no path from node i to node j. Simple generalisations of this definition can be studied, e.g., we could consider weighted edges. The graph surface function, $S(r)$, is defined as the number of nodes which are exactly at a distance r from a given node, averaged over all nodes of the network. In what follows, we always start with a finite graph and consider the limit as the size increases indefinitely. The complex network zeta function $\zeta_G(\alpha)$ is defined as

$$\zeta_G(\alpha) := \frac{1}{N}\sum_i \sum_{j\neq i} r_{ij}^{-\alpha}, \tag{1}$$

where N is the graph size, measured by the number of nodes. When $\alpha \to 0$ the number of nodes contributing finite amounts to the sum in Eq. 1 is unbounded, so the sum diverges (when taking the infinite size limit we first take $\alpha \to 0$ and then $N \to \infty$, so $r_{ij}^{-\alpha} \to 1$). When we start with the finite graph, $\zeta_G(0)$ is $N-1$, and it diverges when $N \to \infty$. When the exponent α tends to infinity, the sum in Eq. 1 gets contributions only from the nearest neighbours of a node. The other terms tend to zero. Thus, $\zeta_G(\alpha)$ tends to the average degree $\langle k \rangle$ for the graph as $\alpha \to \infty$. The average degree $\langle k \rangle$ is related to the number of edges e in the graph by $\langle k \rangle = 2e/N$. Thus,

$$\langle k \rangle = \lim_{\alpha\to\infty} \zeta_G(\alpha) = 2e/N. \tag{2}$$

The definition Eq. 1 can be expressed as a weighted sum over the node distances. This gives the Dirichlet series relation

$$\zeta_G(\alpha) = \sum_r S(r)/r^\alpha. \tag{3}$$

$\zeta_G(\alpha)$ is a decreasing function of α, $\zeta_G(\alpha_1) > \zeta_G(\alpha_2)$, if $\alpha_1 < \alpha_2$. Since the average degree of the nodes ($\langle k \rangle$) is finite, as N goes to ∞ there is exactly one value of α, $\alpha_{transition}$, below which $\limsup_{N\to\infty} \zeta_G(\alpha)$ is infinite and above which it is finite. This has been defined as the dimension of the complex network [14]. For regular discrete d-dimensional lattices $\mathbf{Z}^d$ with distance defined using the L^1 norm

$$\|\vec{n}\|_1 = |n_1| + \cdots + |n_d|, \tag{4}$$

the transition occurs at $\alpha = d$. This is because for a discrete regular lattice of dimension d $S(r)$ grows asyptotically as r^{d-1}, and hence the sum in the complex network zeta function (Eq. 3) converges for $\alpha > d$.

The dimension definition can be related to the Kolmogorov complexity of the graph [10] as defined in computational complexity theory. To fix the notation, we recall that a graph of size N can be represented by specifying the presence or absence of the $N(N-1)/2$ possible edges. Thus, it can be specified as a binary string E of length $N(N-1)/2$. The definition of the randomness deficiency $\delta(N)$ of a graph (Definition 6.4.2 of [10]) is

$$C(E|N,\delta) \geq N(N-1)/2 - \delta(N) \tag{5}$$

where $C(E|N,\delta)$ is the conditional Kolmogorov complexity of E.

Theorem 1 *A graph with finite dimension will have a large randomness deficiency, $\delta(N) = \Omega(N)$.*

PROOF. Lemma 6.4.2 of Ref. [10] states that the degree k of each node of the graph satisfies

$$|k-(N-1)/2| = O\left(\sqrt{(\delta(N)+\log(N))N}\right). \tag{6}$$

The dimension of the graph being finite implies that $\langle k \rangle$ is finite. Thus, Eq. 6 can be valid only if $\delta(N) = \Omega(N)$. $\square$

3 Node Based definition

The definition of dimension based on Eq. 3 involves considering a class of finite large graphs, and taking the limit as the size of the graphs in the class tends to ∞. It would be desirable to define the dimension without taking averages over the nodes of the graph, e.g., for the graph surface function. In this section we show that one can do so, by defining the dimension for a single node, and defining the dimension for the graph as the supremum of the node dimensions.

For any given node i, let us define the node surface function $S_i(r)$ as the number of nodes which are exactly at a distance r from the given node, and the node zeta function $\zeta_i(\alpha)$ as

$$\zeta_i(\alpha) := \sum_{j\neq i} r_{ij}^{-\alpha}. \tag{7}$$

When $\alpha \to 0$ the number of nodes contributing finite amounts to the sum in Eq. 7 is unbounded, so the sum diverges. When the exponent α tends to infinity, the sum in Eq. 7 gets contributions only from the nearest neighbours of the node. Thus, $\zeta_i(\alpha)$ tends to the number of nearest neighbours of the node as $\alpha \to \infty$. The definition Eq. 7 can be expressed as a weighted sum over the node distances. This gives the Dirichlet series relation

$$\zeta_i(\alpha) = \sum_r S_i(r)/r^\alpha. \tag{8}$$

$\zeta_i(\alpha)$ is a decreasing function of α, $\zeta_i(\alpha_1) > \zeta_i(\alpha_2)$, if $\alpha_1 < \alpha_2$. Since we require that the number of nearest neighbours for every node be bounded, then there is exactly one value of α, $\alpha_{transition,i}$, at which the node zeta function transitions from being infinite to being finite. The dimension of the complex network can be defined as

$$\limsup_i \alpha_{transition,i}. \tag{9}$$

This defintion of the dimension is not identical with the definition based on Eq. 3, but its behaviour is similar to the earlier definition, and it is less sensitive to the details of how the infinite size limit is approached.

The properties of dimension like monotonicity (a subset has a lower or the same dimension as its containing set), stability (a union of sets has the maximum dimension of the component sets forming the union) and Lipschitz invariance [7] depend on the details of the set operations. For example, if the number of common edges and nodes in the sets whose union is being taken is small compared to the total number of edges and nodes respectively, then these properties are satisfied.

4 Entropy Characterization

In the theory of effective fractal dimensions [1,5,6,11–13], one can relate the dimension to entropy rates [16,17] for one-sided infinite sequences. It is interesting that one can derive a similar relation for the dimension of complex networks as defined in section 2, with the path counts taking the place of the entropy rates that occur in computer science applications. We will derive the result in this section. It is essentially an application of Cahen's result for the convergence of Dirichlet's series [4]. Let us define volume for a complex network (see [15]) as

$$V(r) = \sum_{i=1}^{r} S(i). \tag{10}$$

The analogue of entropy rate can then be defined as

$$\gamma = \limsup_{r\to\infty} \frac{\log V(r)}{\log r}. \tag{11}$$

We will show that this is equal to the dimension defined in section 2. In the proof one needs to use Abel's lemma on partial summation [8]. The following notation will be needed [8]. Let a_n represent a decreasing sequence. We define the following:

- $A(x) = \sum_{n=1}^{x} a_n$,
- $A(x, y) = \sum_{n=x}^{y} a_n$,
- $\Delta a_n = a_n - a_{n+1}$.

Lemma 2 *(Abel's lemma)* $\sum_{n=x}^{y} a_n b_n = \sum_{n=x}^{y-1} A(x, n)\Delta b_n + A(x, y)b_y$, *for any decreasing sequence* b_n.

In terms of the functions we are studying, Abel's lemma can be written as

$$\sum_{r=1}^{y} S(r)r^{-\alpha} = \sum_{r=1}^{y-1} V(r)\Delta(r^{-\alpha}) + V(y)y^{-\alpha}. \tag{12}$$

The result of this section can be stated as the following theorem.

Theorem 3 *The transition of the series 3 from non-convergence to convergence occurs when α crosses the value γ defined in Equation 11 from below, i.e., $\alpha_{transition} = \gamma$.*

PROOF. The proof is in two parts. We first prove that the series 3 converges for $\alpha = \gamma+\delta$, where δ is any positive number. Let us choose ϵ such that $0 < \epsilon < \delta$. From Equation 11, for n large enough we have $\log V(r) < (\gamma + \delta - \epsilon) \log r$, i.e., $V(r) < r^{\gamma+\delta-\epsilon}$. From Abel's summation lemma,

$$\sum_{r=1}^{n} S(r) r^{-\alpha} = \sum_{r=1}^{n-1} V(r) \Delta r^{-\alpha} + V(n) n^{-\alpha}. \tag{13}$$

The second term goes to zero as $n \to \infty$, therefore we only need to prove the convergence of the first term. By the definition of γ,

$$\sum_{r=1}^{n-1} V(r) \Delta r^{-\alpha} < \sum_{r=1}^{n-1} r^{\gamma+\delta-\epsilon} \Delta r^{-(\gamma+\delta)}. \tag{14}$$

Since $(\gamma + \delta - \epsilon)$ is positive, the term in the sum on the right hand side can be written as

$$r^{\gamma+\delta-\epsilon} \Delta r^{-(\gamma+\delta)} = (\gamma + \delta) \int_{\log r}^{\log(r+1)} e^{(\gamma+\delta-\epsilon)\log r - (\gamma+\delta)x} dx \tag{15}$$

which is less than $(\gamma+\delta) \int_{\log r}^{\log(r+1)} e^{-\epsilon x} dx$, and the series $(\gamma+\delta) \sum_{r=1}^{n-1} \int_{\log r}^{\log(r+1)} e^{-\epsilon x} dx$ is obviously convergent. It follows that $\alpha_{transition} \leq \gamma$.

We next prove that if the series 3 converges, then $\alpha \geq \gamma$. Let us consider a value of α for which $\sum S(r) r^{-\alpha} := \sum_{r=1}^{\infty} b(r)$ converges. Then

$$V(r) = \sum_{r=1}^{n} b(r) r^{\alpha} = \sum_{r=1}^{n-1} B(r) \Delta r^{\alpha} + B(n) n^{\alpha}. \tag{16}$$

Since we are considering a value of α for which the sum $\sum_{r=1}^{\infty} b(r)$ converges, $B(n)$ in Eq. 16 is bounded. Thus, for n sufficiently large, we can find a constant K such that $V(r) < K n^{\alpha}$, i.e.,

$$\log V(r) < \alpha \log n + \log K < (\alpha + \delta) \log n, \tag{17}$$

for any positive δ. Hence

$$\alpha \geq \limsup_{n \to \infty} \frac{\log V(n)}{\log n} = \gamma, \tag{18}$$

and therefore $\alpha_{transition} \geq \gamma$.

Thus, we find that $\alpha_{transition} = \gamma$, as stated in the theorem. $\square$

For completeness we mention that one can define a lower dimension by

$$\liminf_{r \to \infty} \frac{\log V(r)}{\log r}. \tag{19}$$

This would be analogous to the box dimension defined in other discrete contexts (see [9] for a discussion of fractal dimensions in the computer science context). Further study of the analogy may be interesting.

5 Connectedness

A theorem in classical fractal geometry states that any set of dimension less than 1 is totally disconnected. This was extended to discrete regular lattices $\mathbf{Z}^d$ by Doty et al [6]. We show that the proof of [6] can be generalized to hold for complex networks. For positive integers d, r, and points $\vec{m}, \vec{n}$ in $\mathbf{Z}^d$, Doty et al [6] define an *r-path* from $\vec{m}$ to $\vec{n}$ as a sequence $\pi = (\vec{p_0}, \ldots, \vec{p_l})$ of points $\vec{p_i} \in Z^d$ such that $\vec{p_0} = \vec{m}$, $\vec{p_l} = \vec{n}$, and $\|\vec{p_i} - \vec{p_{i+1}}\|_d \leq r$ for all $0 \leq i < l$, where the norm is for $\mathbf{Z}^d$ and define a set $A \subseteq \mathbf{Z}^d$ to be *boundedly connected* if there exists a positive integer r such that, for all $\vec{m}, \vec{n} \in A$, there is an r-path $\pi = (\vec{p_0}, \ldots, \vec{p_l})$ from $\vec{m}$ to $\vec{n}$ in which $\vec{p_i} \in A$ for all $0 \leq i \leq l$. We will extend these definitions to complex networks, as detailed below. For a given complex network of dimension α and a positive integer r, and nodes m, n in the graph, we define an *r-path* from m to n as a sequence $\pi = (p_0, \ldots, p_l)$ of nodes p_i belonging to the complex network, such that $p_0 = m$, $p_l = n$, and $\|p_i - p_{i+1}\| \leq r$ for all $0 \leq i < l$, where the norm is for the complex network. We define a subset of the complex network A to be *boundedly connected* if there exists a positive integer r such that, for all $m, n \in A$, there is an r-path $\pi = (p_0, \ldots, p_l)$ from m to n in which $p_i \in A$ for all $0 \leq i \leq l$. Then the following theorem, and its proof, carry over almost unchanged from the Z^d results proved in [6], with appropriate re-interpretation of the meanings of the symbols.

Theorem 4 *Let A be a complex network. If dimension$(A) < 1$, then no infinite subset of A is boundedly connected.*

6 Conclusions

The complex network zeta function and the graph surface function have been used in different applications, including defining the dimension of a large graph or complex network. In this paper we showed the relation to concepts in computer science. We studied the relation to Kolmogorov complexity. We formulated a definition of complex network dimension which uses limit supremum to avoid averaging over all the nodes of the network, thus making the results easier to apply to formally infinite graphs. The functions have interesting similarities to the concepts used in the study of fractal dimension in computer science studies, like complexity classes. For example, zeta dimension is related to the entropy rate of infinite sequences. In searching for an analogous result for complex netwrk dimension we found a new result, that for graphs an analogous role is played by the count of paths and the growth of the count with the path length. Another new result that was shown by the analogy is the relation between dimension and connectedness. It would be interesting to pursue the analogies further. For example, it would be useful if one can find a s-gale [10] characterization of complex network dimension, since the similar characterization for zeta dimension reveals many properties of the dimension.

References

[1] K.B. Athreya, J.M. Hitchcock, J.H. Lutz and E. Mayordomo. Effective strong dimension in algorithmic information and computational complexity *SIAM Journal on Computing, to Appear*(2007).

[2] T. M. Apostol. *Modular Functions and Dirichlet Series in Number Theory*, in *Graduate Texts in Mathematics, 2nd ed.* (Springer, 1990)**41**.

[3] T. M. Apostol. *Introduction to Analytic Number Theory*, *Undergraduate Texts in Mathematics* (Springer-Verlag, 1976).

[4] E. Cahen. Sur la fonction $\zeta(s)$ de Riemann et sur des fonctions analogues. *Annales de l'École Normale supérieure*, **3** (1894), 75-164.

[5] J. J. Dai, J. I. Lathrop, J. H. Lutz, and E. Mayordomo. Finite-state dimension, *Theoretical Computer Science* **310**(2004),1-33.

[6] D. Doty, X. Gu, J. H. Lutz, E. Mayordomo, and P. Moser, Zeta-dimension, in *Proceedings of the 30th International Symposium on Mathematical Foundations of Computer Science*, (Springer-Verlag, 2005), pp.283-294.

[7] K. Falconer. *Fractal Geometry: Mathematical Foundations and Applications.* Wiley, second edition, 2003.

[8] G. H. Hardy and Marcel Riesz, *The General Theory of Dirichlet's Series*, (Cambridge University Press, Cambride, 1915).

[9] John M. Hitchcock. *Effective Fractal Dimension: Foundations and Applications*, (Ph. D. Thesis, Iowa State University, 2003).

[10] M. Li and P. M. B. Vitányi. *An Introduction to Kolmogorov Complexity and its Applications*, (Springer-Verlag, Berlin, 1997, Second Edition).

[11] J. H. Lutz. Dimension in complexity classes. *SIAM Journal on Computing* **32**(2003),1236-1259.

[12] J. H. Lutz. The dimensions of individual strings and sequences. *Information and Computation* **187**(2003),49-79.

[13] J. H. Lutz. Effective fractal dimensions. *Mathematical Logic Quarterly* **51**(2005),62–72.

[14] O. Shanker, Zeta Function and Dimension of Complex Network, *Mod. Phys. Lett.* **B21**(2007), 639-644.

[15] O. Shanker, Defining Dimension of a Complex Network, *Mod. Phys. Lett.* **B21**(2007), 321-326.

[16] L. Staiger. Kolmogorov complexity and Hausdorff dimension, *Information and Computation*, **103**(1993)159-194.

[17] L. Staiger. A tight upper bound on Kolmogorov complexity and uniformly optimum prediction, *Theory of Computing systems*, **31**(1998)215-229.