Estimating the Manifold Dimension of a Complex Network Using Weyl's Law

Luca Rossi[1] and Andrea Torsello[2]

[1] Queen Mary University of London, London, United Kingdom
luca.rossi@qmul.ac.uk
[2] Università Ca' Foscari Venezia, Venice, Italy
andrea.torsello@unive.it

Abstract. The dimension of the space underlying real-world networks has been shown to strongly influence the networks structural properties, from the degree distribution to the way the networks respond to diffusion and percolation processes. In this paper we propose a way to estimate the dimension of the manifold underlying a network that is based on Weyl's law, a mathematical result that describes the asymptotic behaviour of the eigenvalues of the graph Laplacian. For the case of manifold graphs, the dimension we estimate is equivalent to the fractal dimension of the network, a measure of structural self-similarity. Through an extensive set of experiments on both synthetic and real-world networks we show that our approach is able to correctly estimate the manifold dimension. We compare this with alternative methods to compute the fractal dimension and we show that our approach yields a better estimate on both synthetic and real-world examples.

Keywords: Manifold dimension · Complex networks · Weyl law

1 Introduction

Graphs have long been used as natural representations for a variety of real-world systems, from biological systems [7] to transportation networks [6] and human interactions [10, 2]. These graphs often display non-trivial topological features and are hence referred to as complex networks. The ultimate goal when analysing these networks is that of establishing a link between the structural properties of the networks and their function. To this end, a large number of techniques, from node centralities [13, 9] to entropy measures [11], have been introduced to capture the local and global structural properties of a network [18, 17, 15, 14].

Many real-world networks are embedded in either a two-dimensional or a three-dimensional space, such as the network of collaborations between software developers across the world [10] or railway networks [6]. It has been shown that the structural properties of these networks are strongly influenced by the geometry of the underlying space. In the case where the underlying space is hyperbolic, Krioukov et al. [8] have shown that heterogeneous degree distributions and strong clustering naturally emerge as consequences of the negative curvature and metric property of the space. When the underlying space is Euclidean, the network is often referred to as a spatially embedded network and it's been observed that the probability of establishing a connection between two nodes tends to decay exponentially as the distance between them increases [5].

Daqing et al. [5] proposed a way to measure the dimension of spatially embedded networks. This is achieved under the assumption that the Euclidean distance between the nodes is known and it requires measuring the average distance between the nodes of subgraphs of increasing radius centered around randomly chosen seed nodes. What Daqing et al. compute is effectively the fractal dimension of the network [18], a measure of the self-similarity of the network structure. It's easy to show that the fractal dimension of a network is equivalent to the dimension of the embedding space on regular lattices or in general manifold graphs, i.e., graphs that can be seen as discrete representations of the continuous underlying manifold. Interestingly, Daqing et al. [5] showed that the dimension of the network is intimately related to the properties of diffusion and percolation processes on the network.

Song et al. [17] explored instead two alternative methods to estimate the fractal dimension of a network. The first method is very similar to [5] and involves repeatedly sampling a set of random nodes in the network which are used to grow clusters of nodes whose size is used to ultimately estimate the fractal dimension of the network. In practice, this approach is shown to perform poorly in networks with inhomogeneous degree distributions. A second method estimates the fractal dimension of a network based on the box covering algorithm. This is however an NP-hard problem so heuristics are needed to find an approximate solution [17].

In this paper, we propose an alternative way to estimate the manifold dimension of a weighted network, where the weights are not restricted to represent Euclidean distances between the nodes as in [5]. Specifically, we propose to estimate the manifold dimension of a network using Weyl's law [21]. In spectral theory, Weyl's law describes the asymptotic behaviour of the eigenvalues of the Laplacian associated to a bounded domain $\Omega \in \mathbb{R}^d$, establishing a power-law relation between the eigenvalues and their indices. Crucially, the exponent of this power-law relation depends on the dimension of the underlying manifold. As a result, given a network, we are able to estimate the dimension of the underlying manifold from the spectrum of its Laplacian.

The remainder of the paper is organised as follows: Section 2 provides a brief overview of the two most commonly used approaches to compute the fractal dimension of a graph. Section 3 reviews Weyl's law and introduces our methodology for estimating the manifold dimension of a network, which is then evaluated on both synthetic and real-world networks in Section 4. Finally, Section 5 concludes the paper.

2 Background

Similarly to the more general concept of fractal dimension of a set, the fractal dimension of a network tells us something about how the structure of the network changes as we view it under lenses of varying size. In other words, the fractal dimension of a network is a measure of how invariant or self-similar a network is under a length-scale transformation [18].

Existing approaches to estimate the fractal dimension of a network are based either on the box counting method or the cluster growing method. For a given network G and box size l_B, the box counting method (also known as box covering method) defines a box as a set of nodes such that the distance between any two nodes in the set is smaller than l_B. The number of boxes of size l_B required to cover the network is $N_B(l_B)$ and the goal of the box counting method is to find the minimum value of $N_B(l_B)$ for any value of l_B. As shown in [17], this problem can be mapped to the NP-hard graph colouring problem, so it's typically solved using a number of different heuristics. Given the optimal values of $N_B(l_B)$ for a varying number of box sizes, the fractal dimension d_B of the network is given by

$$N_B(l_B) \approx l_b^{-d_B} \,. \tag{1}$$

Note that, as a consequence of the heuristic nature of the algorithms used to approximate the solution of the box covering problem, the minimum number of boxes for a given size is likely to be overestimated and thus the fractal dimension is instead underestimated.

The cluster growing method instead selects a number of seed nodes at random. For each seed, a cluster is defined as the set of nodes a distance less or equal to l_C from the seed. For each cluster we compute the mass $M_C(l_C)$ as the number of nodes in the cluster. Then the fractional dimension d_C is given by

$$\overline{M_C(l_C)} \approx l_C^{d_C} \,, \tag{2}$$

where $\overline{M_C(l_C)}$ is the average mass of the clusters for a given value of l_C [18]. The main drawback of this approach is that it performs poorly on networks with inhomogeneous degree distributions. This is because by choosing the seeds at random there is a high probability of including hubs in the clusters, leading to a biased estimate of the fractal dimension [17].

3 Weyl's law and the manifold dimension of a network

Let $\Omega \in \mathbb{R}^d$ be a bounded domain and λ_j denote the j-th eigenvalue of the Laplacian on this domain. Weyl's law [21] states that

$$\lim_{\lambda \to \infty} \frac{N(\lambda)}{\lambda^{\frac{d}{2}}} = \frac{\omega_d \mathrm{vol}(\Omega)}{(2\pi)^d} \tag{3}$$

where $N(\lambda) = \#\{\lambda_j \leq \lambda\}$ is a function that counts the number of eigenvalues less than or equal to λ and ω_d is the volume of the unit ball in $\mathbb{R}^d$. Eq. 3 tells us that, for sufficiently large λ,

$$N(\lambda) \approx k\lambda^{\frac{d}{2}} \tag{4}$$

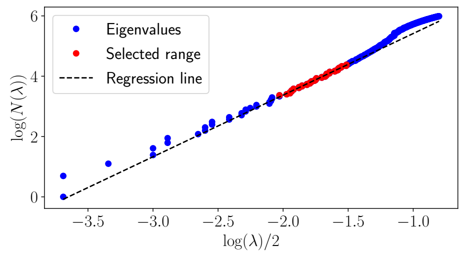


Fig. 1: Estimating the manifold dimension of a two-dimensional 20×20 grid graph. The blue dots correspond to the $(\log N(\lambda_j), \log \lambda_j/2)$ pairs computed on the eigenvalues $\lambda_1, \cdots, \lambda_{400}$ of the graph Laplacian. The manifold dimension is estimated on a selected range of eigenvalues (highlighted in red) to take into account the conditions of Weyl's law and compensate for the boundary effects.

where we used k to group the constants wrt to λ. Taking the logarithm of both sides of Eq. 4 and ignoring the constant term, we get

$$\log N(\lambda) \approx d \frac{\log \lambda}{2} \,. \tag{5}$$

3.1 Estimating the manifold dimension of a network

Let G be a manifold graph, or in other words a graph that accurately models the topology of an underlying manifold of dimension d. Then Eq. 3 holds for the eigenvalues of the Laplacian L of G and we can estimate the dimension d from Eq. 5. Specifically, we use the slope of the regression line on the points $(\log N(\lambda_j), \log \lambda_j/2)$ as an estimate of d.

Fig. 1 shows the values of the function in Eq. 5 sampled on the Laplacian spectrum of a two-dimensional 20×20 grid graph. The slope of the regression line in Fig. 1 is ~ 2, confirming that in this instance our approach is able to accurately capture the graph manifold dimension.

Note that the linear regression is best computed over a selected range of the spectrum (highlighted in red in the toy example of Fig. 1) which excludes the lowest and highest regions. This is because Eq. 3 doesn't hold for low frequencies and high frequencies end up capturing the local variations of the dimension near the graph boundary.

Dealing with edge weights and node attributes Our approach can easily take into account potential information on edge weights and node attributes by incorporating them into the Laplacian. To this end, we turn distances and dissimilarities between the nodes into similarities through a negative exponential transformation.

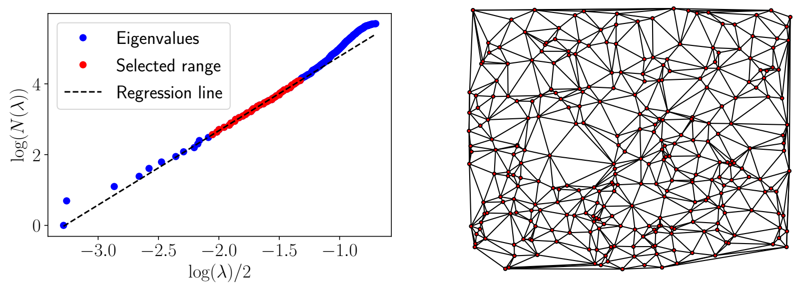


Fig. 2: Sample two-dimensional Delaunay graph (right) over 200 nodes and corresponding log-log plot (left). The slope of the regression line is $d = 2.08$.

4 Experimental evaluation

We perform an extensive set of experiments on both synthetic and real-world networks to evaluate the proposed approach. We compare our results with those obtained using the Maximum-Excluded-Mass-Burning (MEMB) algorithm of Song et al. [17]. Specifically, we used the implementation made available by Akiba et al. [1] at https://github.com/kenkoooo/graph-sketch-fractality. Unfortunately we were unable to find any implementation of MEMB or alternative algorithms that could take edge weights into account. To our understanding, it should be relatively simple to extend MEMB and similar algorithms to deal with networks where the distance between the nodes is available. Indeed Wei et al. discuss such an extension in [20] but fail to provide an implementation of their algorithm. For this reason, in the following experiments when comparing our method with MEMB we show the results both with and without edge weights.

Finally, as discussed in Section 3, our method requires selecting a range of the sorted eigenvalues to sample the values of the function in Eq. 5 and estimate the manifold dimension. Unless otherwise stated, all the experiments in this paper are performed keeping only the eigenvalues in the 7% to 20% range (see Fig. 1).

4.1 Synthetic networks

Delaunay graphs We sampled 200 points uniformly on a two-dimensional plane and we computed their Delaunay triangulation. We repeated this process 100 times and obtained 100 Delaunay graphs. Fig. 2 shows one sample Delaunay graph and the eigenvalues plot computed according to Eq. 5. The weights on the edges of these graphs correspond to the Euclidean distance between the corresponding pair of points. Note that these are manifold graphs embedded on a two-dimensional space, so their manifold dimension is 2.

Table 1: Manifold dimension estimated by our method and MEMB [17] for two-dimensional Delaunay graphs and hypercubes of increasing dimension.

Network	Delaunay (d=2)	Hypercube (d=2)	Hypercube (d=3)	Hypercube (d=4)
Weyl	2.12 ± 0.01	2.00	3.07	4.05
MEMB	1.66 ± 0.01	1.32	1.83	2.16

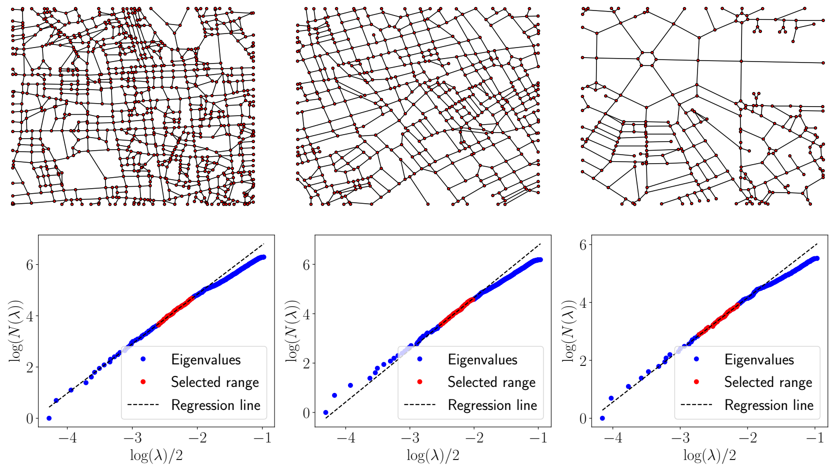
Table 2: Manifold dimension estimated by our method and MEMB [17] on (u, v)-flower networks with increasing fractal dimension.

Network	$(2, 2)$-flower	$(2, 3)$-flower	$(2, 4)$-flower	$(2, 5)$-flower	$(2, 6)$-flower
Weyl	2.01	1.91	1.74	1.67	1.62
MEMB	1.37	1.58	1.69	1.86	2.01

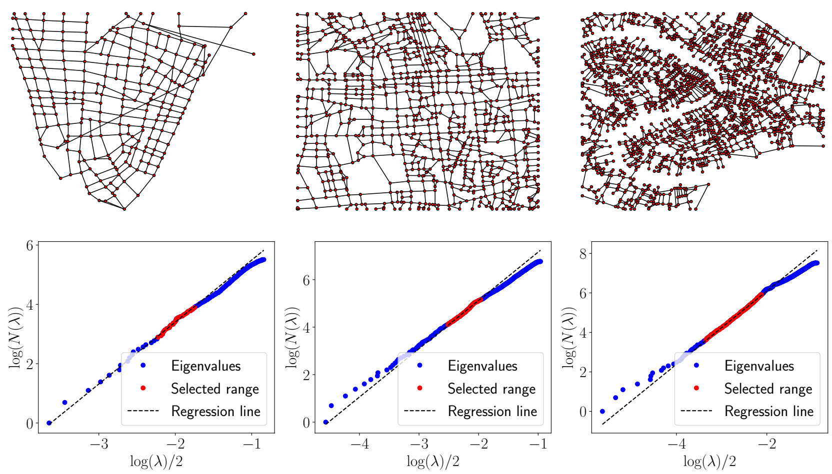
Hypercubes We construct three hypercubes of increasing dimension: 1) one two-dimensional hypercube of side 10, for a total of 100 nodes, 2) one three-dimensional hypercube of side 8, for a total of 512 nodes, and 3) one four-dimensional hypercube of side 6, for a total of 1296 nodes. The manifold dimension of the hypercubes is 2, 3, and 4, respectively.

Table 1 shows the values of the manifold dimension estimated by our method and MEMB [17] on the synthetic datasets. For the Delaunay graphs we report the average value of the dimension $\pm$ standard error. Note that MEMB consistently underestimates the manifold dimension of the graphs. As explained in Section 2, this is because computing the optimal box covering is NP-hard and thus the solution found by heuristic approaches like MEMB is likely to overestimate the number of boxes, leading to an underestimation of the manifold dimension. The value estimated with our method, instead, consistently falls very close to the ground truth. This is true even if we drop the edge weights in the Delaunay graphs. In this case, the average manifold dimension is estimated to be 2.31 ± 0.01.

(u, v)-flowers We also compare our method and MEMB on an additional set of synthetic graphs where the fractal dimension can be computed analytically [16]. Starting from a cycle graph consisting of $u + v$ nodes, new nodes and edges are iteratively added by replacing each edge by two parallel paths, u and v edges long. When $2 \leq u < v$, it can be shown that the network has a finite fractal dimension equal to $\ln(u + v)/\ln(u)$. By fixing $u = 2$ and letting v grow, we can create networks of increasing fractal dimension. While this trend is correctly captured by MEMB, the dimension estimated with our method decreases as v increases, as Table 2 shows. This isn't surprising, as by fixing u and letting v grow we are effectively creating a network that contains increasingly long uni-dimensional string-like structures (see for example Fig. 2a in [16]). Indeed, the flower graphs are fractal but not manifold, so our method cannot be applied.



(a) Bologna (b) London (c) New Delhi



(d) New York (e) Seoul (f) Venice

Fig. 3: Graphs (top) and corresponding log-log plots (bottom) for the urban street networks of Bologna ($d = 1.93$), London ($d = 2.11$), New Delhi ($d = 1.80$), New York ($d = 2.09$), Seoul ($d = 2.03$), and Venice ($d = 2.00$).

4.2 Urban street networks

We consider 6 urban street networks corresponding to 1-square mile maps of Bologna (541 nodes and 771 edges), London (488 nodes and 729 edges), New Delhi (252 nodes and 328 edges), New York (248 nodes and 418 edges), Seoul (869 nodes and 1,307 edges), and Venice (1,840 nodes and 2,397 edges) [4, 9]. The edge weights of these graphs correspond to the length of the road segment connecting

Table 3: Manifold dimension estimated by our method and MEMB [17] on the urban street networks of 6 cites around the world.

Network	Bologna	London	New Delhi	New York	Seoul	Venice
Weyl	1.93 (*2.05*)	2.11 (*2.11*)	1.80 (*1.78*)	2.09 (*2.12*)	2.03 (*2.07*)	2.00 (*2.00*)
MEMB	1.60	1.55	1.56	1.52	1.59	1.50

the two endpoints. Fig. 3 shows the graph of the 6 cities and the corresponding log-log plots. Table 3 shows the value of the manifold dimension estimated by our method and MEMB. In all 6 cases our method gives a result that is significantly closer to what we expect to be ground truth for these networks (2), with New Delhi having the lowest dimension. This in turn may be due to the particular structure of the part of New Delhi captured in this dataset, with large areas covered only by a small number of long roads, effectively lowering the estimated dimensionality. As observed for the Delaunay graphs, removing the edge weights has only a minimal effect on the estimate (italic in Table 3).

4.3 Other real-world networks

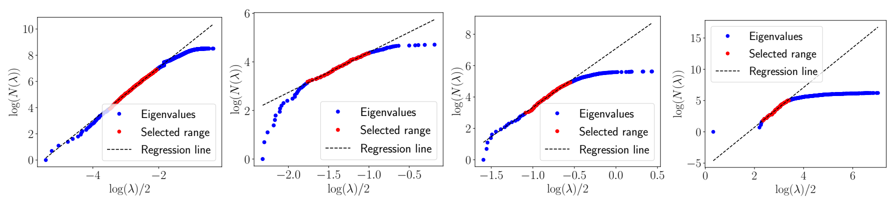
US power grid This network represents the high-voltage power grid of the US (Western states). The nodes (4,941) are transformers, substations, and generators, and the edges (6,594) represent transmission lines [19]. No edge weight information or node coordinates were available for this network. The range of the eigenvalues used to fit the regression line is 1% to 20%.

Dickens This network represents the most commonly used adjectives and nouns in the novel David Copperfield by Charles Dickens. The network has 112 nodes with 425 edges connecting pairs of adjacent words in the text [12]. The edge weights represent the Levenshtein distance between the words. The range of the eigenvalues used to fit the regression line is 2% to 70%.

C. elegans This is an unweighted network representing the Caenorhabditis elegans neuronal network, consisting of 279 nodes representing nonpharyngeal neurons and 2,287 edges representing synaptic connections [19, 9]. The range of the eigenvalues used to fit the regression line is 7% to 50%.

US airports This is the network of flight connections between the 500 US airports with the highest traffic [3, 9]. Each node (500) corresponds to an airport and each edge (2,980) has an integer weight corresponding to the total number of seats available on all the direct routes between the two endpoints within a year. The range of the eigenvalues used to fit the regression line is 1% to 30%.

Fig. 4 shows the log-log plots for these networks and Table 4 lists the estimated manifold dimensions. While *US power grid* and *Dickens* are clearly manifold, this is less obvious for *C. elegans*, where it is harder to distinguish between



(a) Powergrid (b) Dickens (c) C. elegans (d) Airports

Fig. 4: Log-log plots for the Powergrid ($d = 2.13$), Dickens ($d = 1.65$), C. elegans ($d = 3.75$), and Airports networks (3.19).

Table 4: Manifold dimension estimated by our method and MEMB [17] on the Powergrid, Dickens, C. elegans, and Airports networks.

Network	Powergrid	Dickens	C. elegans	Airports
Weyl	2.13	1.65 (*1.70*)	3.75	3.19 (*3.62*)
MEMB	2.34	2.58	2.98	2.93

boundary effects and non-manifold behaviour. The log-log plot for *Airports*, on the other hand, shows at least two separate linear trends, suggesting that this is not a manifold network and thus our approach cannot be applied. In general, note that the more manifold the graph is, the more robust the estimation of the dimension wrt the chosen spectral range is.

5 Conclusion

We proposed a way to estimate the manifold dimension of a network using Weyl's law, a mathematical result that describes the asymptotic behaviour of the eigenvalues of the graph Laplacian. We showed through an extensive set of experiments on both synthetic and real-world networks that our approach is able to correctly estimate the manifold dimension, yielding better estimates than an alternative method based on box counting. Future work will investigate the possibility of automatically selecting the spectral range to fit when estimating the manifold dimension. Having access to larger urban networks, it would also be interesting to see if the local manifold dimension of different subgraphs and communities can be related to other quantities of interest.

References

1. Akiba, T., Nakamura, K., Takaguchi, T.: Fractality of massive graphs: Scalable analysis with sketch-based box-covering algorithm. In: 2016 IEEE 16th International Conference on Data Mining (ICDM). pp. 769–774. IEEE (2016)

2. Chorley, M., Rossi, L., Tyson, G., Williams, M.: Pub crawling at scale: tapping untappd to explore social drinking. In: Proceedings of the International AAAI Conference on Web and Social Media. vol. 10 (2016)
3. Colizza, V., Pastor-Satorras, R., Vespignani, A.: Reaction–diffusion processes and metapopulation models in heterogeneous networks. Nature Physics **3**(4), 276–282 (2007)
4. Crucitti, P., Latora, V., Porta, S.: Centrality measures in spatial networks of urban streets. Physical Review E **73**(3), 036125 (2006)
5. Daqing, L., Kosmidis, K., Bunde, A., Havlin, S.: Dimension of spatially embedded networks. Nature Physics **7**(6), 481–484 (2011)
6. Erath, A., Löchl, M., Axhausen, K.W.: Graph-theoretical analysis of the swiss road and railway networks over time. Networks and Spatial Economics **9**(3), 379–400 (2009)
7. Gursoy, A., Keskin, O., Nussinov, R.: Topological properties of protein interaction networks from a structural perspective (2008)
8. Krioukov, D., Papadopoulos, F., Kitsak, M., Vahdat, A., Boguná, M.: Hyperbolic geometry of complex networks. Physical Review E **82**(3), 036106 (2010)
9. Latora, V., Nicosia, V., Russo, G.: Complex networks: principles, methods and applications. Cambridge University Press (2017)
10. Lima, A., Rossi, L., Musolesi, M.: Coding together at scale: Github asa collaborative social network. In: Proceedings of 8th AAAI International Conference on Weblogs and Social Media (2014)
11. Minello, G., Rossi, L., Torsello, A.: On the von neumann entropy of graphs. Journal of Complex Networks **7**(4), 491–514 (2019)
12. Newman, M.E.: Finding community structure in networks using the eigenvectors of matrices. Physical review E **74**(3), 036104 (2006)
13. Rossi, L., Torsello, A., Hancock, E.R.: Node centrality for continuous-time quantum walks. In: Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR). pp. 103–112. Springer (2014)
14. Rossi, L., Torsello, A., Hancock, E.R.: Measuring graph similarity through continuous-time quantum walks and the quantum jensen-shannon divergence. Physical Review E **91**(2), 022815 (2015)
15. Rossi, L., Torsello, A., Hancock, E.R., Wilson, R.C.: Characterizing graph symmetries through quantum jensen-shannon divergence. Physical Review E **88**(3), 032806 (2013)
16. Rozenfeld, H.D., Havlin, S., Ben-Avraham, D.: Fractal and transfractal recursive scale-free nets. New Journal of Physics **9**(6), 175 (2007)
17. Song, C., Gallos, L.K., Havlin, S., Makse, H.A.: How to calculate the fractal dimension of a complex network: the box covering algorithm. Journal of Statistical Mechanics: Theory and Experiment **2007**(03), P03006 (2007)
18. Song, C., Havlin, S., Makse, H.A.: Self-similarity of complex networks. Nature **433**(7024), 392–395 (2005)
19. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world'networks. nature **393**(6684), 440–442 (1998)
20. Wei, D.J., Liu, Q., Zhang, H.X., Hu, Y., Deng, Y., Mahadevan, S.: Box-covering algorithm for fractal dimension of weighted networks. Scientific reports **3**(1), 1–8 (2013)
21. Weyl, H.: Über die asymptotische verteilung der eigenwerte. Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse **1911**, 110–117 (1911)