Towards Using DFT to Characterize Complex Networks

Danilo R. B. de Araújo, Carmelo J. A. Bastos-Filho and Joaquim F. Martins-Filho

***Abstract*— There are some Network Metrics that are very useful to analyze and model Complex Networks. These metrics, including the spectral-based ones, can be used to retrieve information from the network. As an example, the eigenvalues of the Laplacian matrix can present interesting information about the network topology. We observed that if one applies the Discrete Fourier Transform (*DFT*) over the eigenvalues of the Laplacian Matrix, it is possible to observe different patterns in the *DFT* depending on some properties of the analyzed networks. In this paper, we propose two novel metrics based on the *DFT* samples, named *FZC* and *HVC*, that can be used to identify the type of network. We tested these metrics in networks generated by three different models (Random, Small-World and Scale-free) and in real network benchmarks. The results indicate that one can use the proposed metrics to identify the generational model of the network.**

***Keywords*— Complex Networks, Graph Theory, Network Assessment, Discrete Fourier Transform.**

I. Introduction

Graph theory has been applied to practical problems since its proposition in 1736, when the swiss mathematician Leonhard Euler created the principles of this theory aiming to determine how one should circumnavigate the Bridges of Königsberg. Some of the modern concepts regarding the evolution on graphs were proposed in the 1950s, when Paul Erdos proposed some novel concepts focusing on random graphs [1]. The Erdos-Renyi (*ER*) random graphs had a huge impact on the development of this area. On the other hand, some remarkable models, such as Power Law networks (Scale-Free networks) [2] and Small World networks [3], were just proposed in the 1990s and have been deeply studied since then. These models allow one to analyze and to model complex and dynamic systems in several areas, such as physics, mathematics, computer science, biology, economics, etc. Because of this multi-disciplinary aspect, this theory and its applications has been called as "Network Science" [4]. When the number of nodes is high and some network behaviors arise, this theory is known as Complex Networks.

Complex Networks have become popular mainly due to their capability to represent virtually any network structure of real-world phenomena. Because of this, several investigations have been proposed to represent the network structure and to analyze the topological features of the networks in terms of Network Metrics (*NM*), including the analysis of dynamical changes on the topology over the time. Recently, there are some studies focusing on the relationship between the structure and the dynamics of Complex Networks [5]. *NM* are quite important to obtain a comprehensive understanding about this relationship. Besides, the quantitative description of the networks properties also provides fundamental subsidies for classifying Complex Networks.

This paper aims to provide some novel ideas on how to use Discrete Fourier Transform (*DFT*) over the eigenvalues of the Laplacian Matrix in order to characterize Complex Networks. The remainder of this paper is organized as follows: Section II provides a brief review on previously proposed metrics to characterize Complex Networks, *i.e.* we show the definition of the most used *NM*; Section III presents two novel *NM*, named *FZC* and *HVC*, to retrieve information of Complex Networks based on *DFT* spectra and some preliminary results; Section IV presents the conclusions.

Danilo R. B. de Araújo and Joaquim F. Martins-Filho¸ Department of Electrical Engineering and Systems, Federal University of Pernambuco, Recife-PE, Brazil, E-mail: jfmf@ufpe.br. Carmelo J. A. Bastos-Filho, University of Pernambuco, Recife-PE, Brazil E-mail: carmelofilho@poli.br

II. A Brief Review on Complex Network Characterization

The analysis of relevant topological properties is one of the major objectives that drives the research on Complex Networks. A topological property of a network is inherently related to the graph that represent this network. This means that all networks with the same topological properties define a family of graphs. Some *NM* have been proposed to quantify relevant topological properties of Complex Networks [4]. This section presents a brief review of the most used *NM*. Several surveys [6], [7] are available and can be used for further studies on the concepts briefly presented in this section.

We consider a Complex Network as a graph $G = (\mathcal{N}, \mathcal{L})$, in which $\mathcal{N}$ and $\mathcal{L}$ denote the set of vertices and the set of edges, respectively. In this paper we just considered unweighted and undirected graphs. Besides, a graph cannot contain self-loops (connections beginning and ending at the same node). We can also define the amount of nodes and links in a network as $N = |\mathcal{N}|$ and $l = |\mathcal{L}|$, respectively.

A. *Structural-based* CNM

The *link density* (q) of a network is defined as the ratio between the number of links that actually exists and the maximum number of links that could exist in the network (*i.e.* if all nodes are connected to all others by a direct link). The *node degree* (D) describes the number of links or neighbor nodes of a given node. The *node degree distribution* defines the probability, $Pr(D)$, of a randomly selected node to have a certain degree D. The average number of links that are connected to a node is called the *average node degree*. Another metric that quantifies the correlation between pairs of nodes is the *assortativity coefficient* (r), in which r lies in the interval $-1 \leq r \leq 1$. When $r > 0$, the network tends to have nodes that are connected to other nodes that present a similar D [8].

The *shortest path* describes the number of hops between a given pair of nodes. The distance distribution is the probability, $Pr(H)$, that a randomly selected pair of nodes presents a shortest path with value equal to H. The longest shortest path (H_{max}) between any pair of nodes is referred as the *diameter* of a graph, which is defined as *diam(G)*.

The *clustering coefficient* (c_i) is the ratio between the number of triangles that contain node i and the number of triangles that could possibly exist if all neighbors of i were interconnected [3]. The *clustering coefficient for the entire graph* c_G is the average of the clustering coefficients of all the network nodes. Several *NM* based on centrality were also defined. As an example, *Betweenness* measures the centrality of a node or link within a graph. Nodes (links) that appear frequently on shortest paths between node pairs present higher betweenness than the ones that rarely appear [9]. *Average node (link) Betweenness* is the average value of the node (link) betweenness over all nodes (links).

A graph is said to be connected if there exists a path between each pair of nodes. When there is no path between at least one pair of nodes, a network is defined as *disconnected*. If there is a link between every pair of nodes in a graph, the graph is defined as *complete* (K_N). The *link connectivity* k_l is the minimum number of links to be removed in order to turn a graph disconnected. The *node connectivity* k_N is defined analogously [10].

B. Spectral-based CNM

This section summarizes the main metrics related to spectral measurements on graphs. In this case, the metrics are closely related to the eigenvalues of the matrices that represent the network. A survey on graph spectra can be found in [6].

In the graph theory, a network can be represented by its *Adjacency matrix (A)*, the *Node Degree matrix (D)* or the *Laplacian matrix (L)*. The *Adjacency matrix* (A) of an undirected graph with N nodes is a N x N matrix, in which the non-diagonal entries (i, j) are equal to "1" if the nodes i and j are adjacent (connected), or "0" otherwise. In *A*, the entries (i, i) are always equal to "0", since we are considering that a node can not be connected to itself. Since we are considering undirected graphs, the *adjacency matrix* is a symmetric matrix. A diagonal matrix, which contains information about the degrees of the nodes, is named *Node Degree matrix* (D). D is used together with the *Adjacency matrix* to build the *Laplacian matrix* (L) of a graph. L is defined as $L = D - A$, in which the non-diagonal entries (i, j) are either "-1" or "0", depending on whether nodes i and j are adjacent or not, respectively, and the diagonal entries (i, i) are equal to the degree of the nodes D_i.

The study on the relationship between a graph and its eigenvalues (and eigenvectors) is referred in the literature as spectral graph theory. All eigenvalues are real for the *Adjacency matrix* [11], whereas all eigenvalues are real and nonnegative for the *Laplacian matrix* [12]. The ordered set of N eigenvalues of *A* or *L* is called the spectrum of the matrix. If there are two graphs with similar sets of eigenvalues, this means that they probably present similar graph structures or graph isomorphism [13]. Based on this, one can suggest to use the eigenvalues of a characteristic matrix of a graph to retrieve information or to classify Complex Networks.

The largest eigenvalue of *A* is denoted as the *spectral radius* (ρ). It could be used for many applications, such as modeling virus propagation in networks [14]. The second smallest eigenvalue of *L* is denoted as the *algebraic connectivity* (λ_{N-1}). A graph is disconnected if $\lambda_{N-1} = 0$. Moreover, if $\lambda_{N-i+1} = 0$ and $\lambda_{N-i} \neq 0$, then a graph has exactly i components. This also means that the multiplicity of zeros in the eigenvalues of the Laplacian matrix corresponds to the number of independent components of the graph. The algebraic connectivity also measures the connectivity of a graph, *i.e.* a higher value for λ_{N-1} implies in a higher difficulty to cut a graph in two independent components [15]. Besides, there are many real-world problems in which λ_{N-1} has an important meaning [12], [16]. Another important spectral metric, called *natural connectivity NC*, aims to characterize the redundancy of alternative routes in a network. *NC* is calculated by quantifying the number of closed chains of all lengths in the network [17]. Both λ_{N-1} and *NC* are commonly used to measure the robustness of real-world networks.

Previous works have also investigated the *density function* of the eigenvalues $f_\lambda(t)$, which is more suitable to analyze the eigenvalues $\{\lambda_m\}_{1 \leq m \leq N}$ in large graphs [18]. The density function is defined by $f_\lambda(t) = \frac{1}{N} \sum_{m=1}^{N} \delta(t - \lambda_m)$, where $\delta(t)$ is the Dirac function [6]. It is possible to recognize specific families of complex networks by analyzing the *density function* of A.

III. Two Novel Spectral Metrics for Assessing Complex Networks

Although the eigenvalues of the characteristics matrices present a lot of useful information, the most simple and used metrics do not involve all of them. In order to better understand the complete set of eigenvalues of the Laplacian matrix of a Complex Network, we propose here to analyze the spectral behavior of the eigenvalues by applying the Discrete Fourier Transform (*DFT*) [19] over the entire set of eigenvalues. Figures 1 and 2 show the Even and Odd components for the *DFT* of the eigenvalues of the *Laplacian matrix* for a scale-free network generated by the generative procedure of Barabási (Scale-Free) [2]. First of all, one can observe that the Even and Odd components for the *DFT* present different behaviors and must be observed in separate. This indicate that there must be a phase coherence in the eigenvalues. We have observed that the *DFT* present different behaviors for other Complex Network models, such as Erdos-Renyi (Random Graphs) [1] and Watts-Strogatz (Small-World) [3].

Based on the characteristics observed in Figures 1 and 2, we propose two simple metrics involving the *DFT* samples. The first metric, namely *FZC*, represents the frequency in which the first even component of the *DFT* crosses the zero. The second

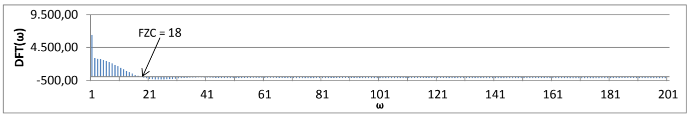


Fig. 1. Even DTF values of the eigenvalues of the *Laplacian matrix* of a scale-free network.

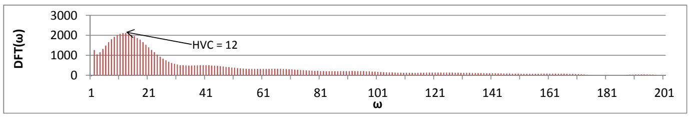


Fig. 2. Odd DTF values of the eigenvalues of the *Laplacian matrix* of a scale-free network.

metric, called *HVC*, represents the frequency in which the odd component presents the highest *DFT* value. The procedures to calculate both metrics are shown in the pseudocode depicted in Algorithm 1.

Algorithm 1: The algorithm used to calculate *FZC* and *HVC*.

Let A_N the adjacency matrix of the network N;
Let D_N the degree matrix of the network N;
Calculate the Laplacian matrix $L_N = D_N - A_N$;
Calculate the real eigenvalues of L_N and store it in E_N;
Calculate the DFT of E_N and store the values in DFT_N;
Let DFT_E be the set of even components of DFT;
Let DFT_O be the set of odd components of DFT;
Let FZC be the first component of the DFT_E that passes through the zero;
Let HVC be the component associated to the highest DFT_O value;
Let $lastValue = DFT_E(0)$;
for *(int c = 0; c < DFT_E.length; c++)* **do**
  if *($abs(signum(DFT_E(c)) + signum(lastValue))$!= 2)* **then**
    FZC = c;
    $break$;
  end
  $lastValue = DFT_E(c)$;
end
Let $maxValue = -\infty$;
for *(int c = 0; c < DFT_O.length; c++)* **do**
  if *($DFT_O(c) > maxValue$)* **then**
    $maxValue = DFT_O(c)$;
    HVC = c;
  end
end

In Algorithm 1, the function $abs(.)$ means that $abs(z) = |z| = |-z| = z$. The function $signum$ is defined as follows:

$$signum(z) = \begin{cases} 1 & \text{for} \quad z \geq 1 \\ -1 & \text{for} \quad z \leq -1 \\ 0 & \text{for} \quad z = 0 \end{cases} . \tag{1}$$

We aim to show that the *DTF* calculated over the eigenvalues of the *Laplacian matrix* present different characteristics for different types of networks. In order to show this, we generated $45,000$ different networks using the generative procedures of Erdos-Renyi (Random), Barabási (Scale-Free) and Watts-Strogatz (Small-World). We varied the density values q from 0.02 to 1.00 with a step value of 0.02. For each pair (generative procedure, q), we created 100 different networks with different sizes, *i.e.* 100, 200 and 400 nodes. Our implementation of Random Graphs establishes a link between a pair (i, j) if a uniform random variable assumes a value below a probability value p. In order to generate Small-World networks we create a k-regular graph and change existing links (i, j) to a new one (k, l) considering a rewiring probability $p = 0.1$. We use the value of density (q) to calculate a value to k. Finally, to generate a Scale-Free network we used the preferential attachment process. The networks starts with $n = 3$ nodes and each of the $(N - 3)$ remaining nodes are attached to the network by adding Δm links to the existing nodes. We use the density (q) to determine the value of Δm.

Figures 3 (top), 4 (top) and 5 (top) show the average value of the *FZC* obtained from 100 different networks with 100, 200 and 400 nodes, respectively, as a function of the density of the network. All networks were generated independently by its referred model. One can observe clearly that the *FZC* occurs for higher frequencies for networks with low density and diminishes as the density increases for all models. Furthermore, Scale-Free networks exhibits highest values of *FZC* when compared to Random and Small-World networks with the same density.

Figures 3 (bottom), 4 (bottom) and 5 (bottom) show the average value of the *HVC* obtained from 100 different networks with 100, 200 and 400 nodes, respectively, as a function of the density of the network. One can observe that Random and Small-World networks have $HVC = 1$. For Scale-Free networks, $HVC = 1$ only occurs for networks with density above 0.20, regardless of the number of network nodes. One can also observe that there is a peak that slightly shifts its position depending on the number of network nodes.

In order to analyze the variation in the values of the metrics to a set of 100 generated networks from each type, we generated some box-plot charts for the density $q = 0.02$. According to Figure 6, one can observe that the metrics present different values according to the type of network, specially for the *FZC*. In Figure 6, SF, SW and RND mean Scale-Free, Small-World and Random, respectively.

We have also analyzed the proposed metrics for two real networks. The first network is a social network that describes the face-to-face behavior of people during a conference [20]. This network presents both the Scale-Free and Small-World effect. The second network is the "Caenorhabditis elegans metabolic network" [21] (a typical Scale-Free network). To emphasize the expressivity of using DFT to characterize complex networks we generated two random networks using the Erdos-Renyi generative procedure with similar size and density. The HVC value was equal to "1" for all networks. The FZC values for these four networks are summarized in Table I.

Figures 8 and 9 show the eigenvalues and the odd and even

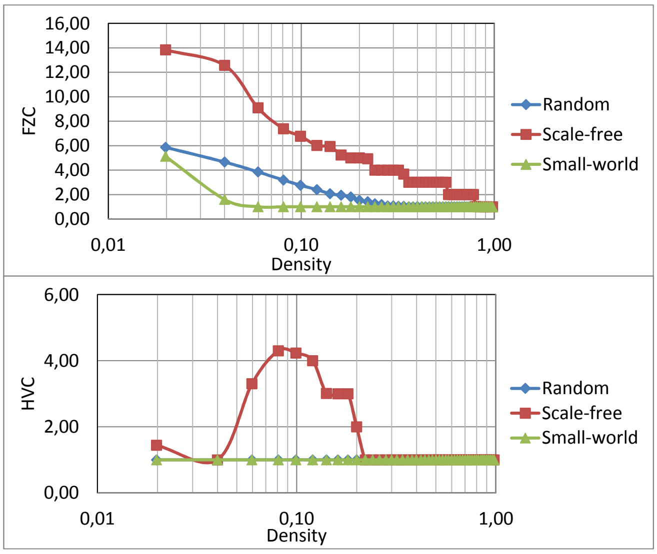


Fig. 3. Comparison of *FZC* (top) and *HVC* (bottom) from networks with 100 nodes as a function of the density.

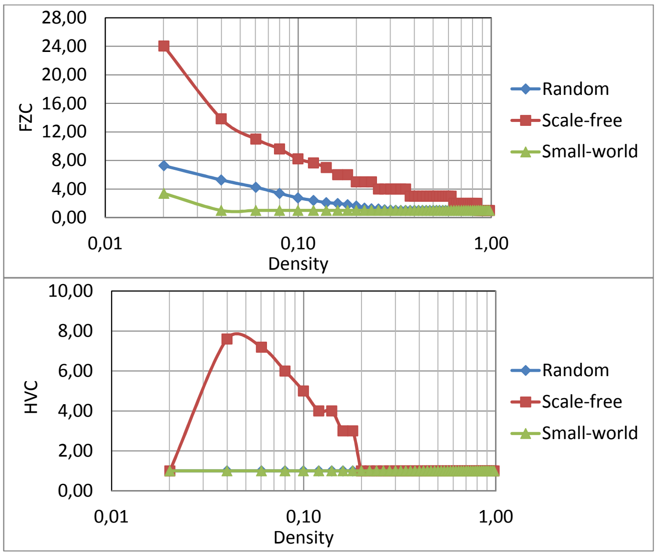


Fig. 4. Comparison of *FZC* (top) and *HVC* (bottom) from networks with 200 nodes as a function of the density.

components of the *DFT* for the two real networks described in Table 1. Figure 7 presents the eigenvalues related with the 3rd network of the Table I. We show the values for the metrics as well. First of all, the DFT curves for each network present different behaviors. From Figures 7(b), 8(b) and 9(b) one can observe that all networks present $HVC = 1$. It is in line with what is shown in Figure 5. Although there is a mismatch for the Scale-Free network ("Caenorhabditis elegans metabolic"), it is possible to observe from Figure 9(b) that the curve presents a similar shape when compared to Figure 2 and the second peak occurs around frequency value "8". Besides, Figures 7(c), 8(c) and 9(c) confirm what it is presented in Figure 5 (top), *i.e.* scale-free networks present the highest FZC, and random networks present a higher FZC value when compared to small-world networks.

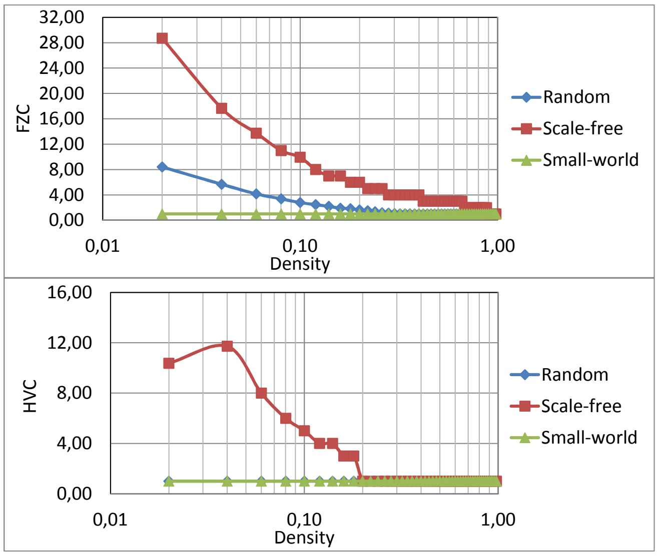
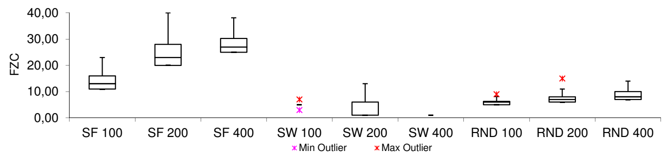
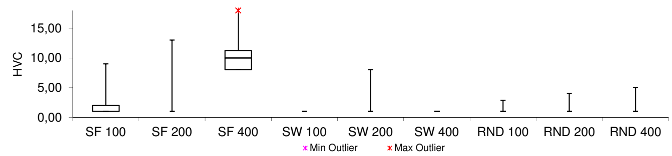


Fig. 5. Comparison of *FZC* (top) and *HVC* (bottom) from networks with 400 nodes as a function of the density.

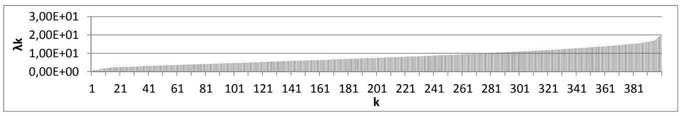


(a) Variation in *FZC* value.

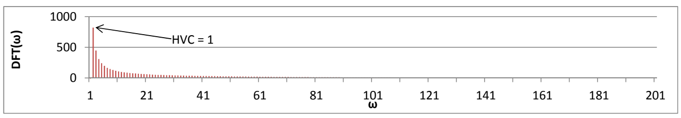


(b) Variation in *HVC* value.

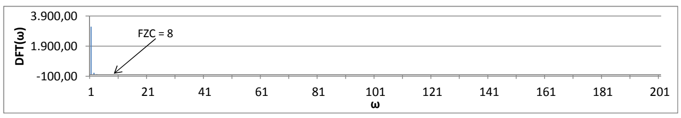
Fig. 6. Box-plot charts to the proposed metrics for networks with density $q = 0.02$.



(a) Eigenvalues of the Laplacian matrix



(b) Odd components of the DFT



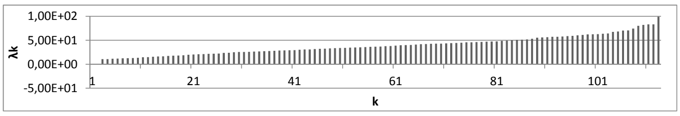
(c) Even components of the DFT

Fig. 7. Eigenvalues of the Laplacian matrix and the DFT of the eigenvalues for a Erdos-Renyi network equivalent to the "Caenorhabditis elegans metabolic" network.

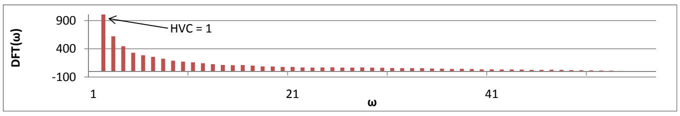
TABLE I

COMPARISON OF TWO DIFFERENT REAL NETWORKS AND AN EQUIVALENT RANDOM NETWORK.

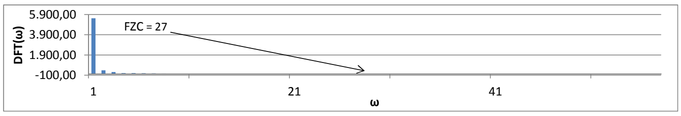
Network	Category	Size	Density	Algebraic Connectivity	FZC
Erdos-Renyi network	Random	113	0.35	22.66	1
Sociopatterns Hypertext 2009	Social	113	0.35	1.00	27
Erdos-Renyi network	Random	453	0.02	1.24	8
C. elegans metabolic	Interaction	453	0.02	0.26	56



(a) Eigenvalues of the Laplacian matrix

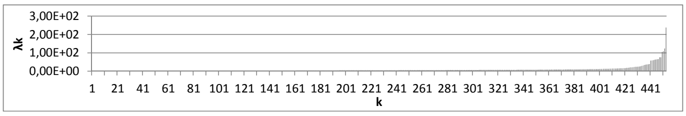


(b) Odd components of the DFT

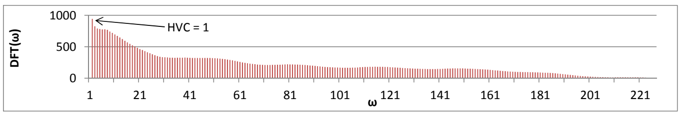


(c) Even components of the DFT

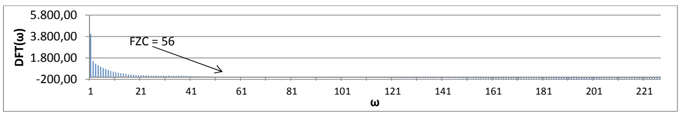
Fig. 8. Eigenvalues of the Laplacian matrix and the DFT of the eigenvalues for the Sociopatterns-Hypertext 2009 network [20].



(a) Eigenvalues of the Laplacian matrix



(b) Odd components of the DFT



(c) Even components of the DFT

Fig. 9. Eigenvalues of the Laplacian matrix and the DFT of the eigenvalues for the "Caenorhabditis elegans metabolic" network [21].

IV. CONCLUSIONS

In this paper we presented some preliminary results towards the use of *DFT* over the eigenvalues of the *Laplacian Matrix* in order to retrieve information or classify different families of complex networks. The inspection of specific points of the DFT curve allows one to correctly separate Scale-Free from Small-World and Random Networks. We also observed that the Even and Odd *DFT* components present different behaviors and, based on this, it is possible to define novel metrics. We propose two metrics, *FZC* and *HVC*, which are related to the even and odd components of the *DFT* over the Laplacian matrix. Both metrics were applied to classify a large group of networks created by the generative procedures of Erdos-Renyi, Albert-Barabási and Watts-Strogatz. We also applied our metrics to real networks in order to assess the results obtained for the simulated networks. Further analysis aims to investigate other characteristics of the *DFT* curves and to establish proper relations between these novel metrics and other well known network metrics.

ACKNOWLEDGEMENTS

The authors acknowledge the financial support from FACEPE, CNPq, CAPES, UFPE and UPE.

REFERÊNCIAS

[1] P. Erdos and A. Rényi. On the evolution of random graphs. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, pages 17–61, 1960.

[2] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[3] D. J. Watts and S. H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393:440442, 1998.

[4] T. G. Lewis. *Network Science - Theory and Applications*. John Wiley & Sons, 2009.

[5] M. E. J. Newman. Structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.

[6] P. van Mieghem. *Graph Spectra for Complex Networks*. Cambridge University Press, 2011.

[7] D. M. Cvetkovic and I. Gutman. *Selected topics on applications of graph spectra*. Beograd : Matematicki institut SANU, 2011.

[8] M. E. J. Newman. Assortative mixing in networks. *Physical Review Letters*, 89:208701, 2002.

[9] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han. Attack vulnerability of complex networks. *Physical Review*, 65:056109, 2002.

[10] B. Bollobás. *Modern graph theory*. Springer-Verlag, 1998.

[11] D. M. Cvetkovic, M. Doob, and H. Sachs. *Spectra of graphs, theory and applications*. Johann Ambrosius Barth Verlag, 1995.

[12] B. Mohar, Y. Alavi, G. Chartrand, O. R. Oellermann, and A. J. Schwenk. The laplacian spectrum of graphs. *Graph Theory, Combinatorics and Applications*, 2:5364, 1991.

[13] E. R. van Dam and W. H. Haemers. Which graphs are determined by their spectrum? *Linear Algebra and its Applications*, 373:241–272, 2003.

[14] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos. Epidemic spreading in real networks: An eigenvalue viewpoint. *Proceedings of the 22nd Symposium in Reliable Distributed Computing*, 2003.

[15] M. Fiedler. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, 23:298305, 1973.

[16] R. Merris. A survey of graph laplacians. *Linear and Multilinear Algebra*, 39:19–31, 1995.

[17] J. Wu, M. Barahona, Y.-J. Tan, and H.-Z. Deng. Spectral measure of structural robustness in complex networks. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 41(6):1244 –1252, nov 2011.

[18] I. J. Farkas, I. Derényi, A. L. Barabási, and T. Vicsek. Spectra of "real-world" graphs: Beyond the semicircle law. *Phys. Rev. E*, 64:026704, Jul 2001.

[19] K. R. Rao, D. N. Kim, and J.-J. Hwang. *Fast Fourier Transform - Algorithms and Applications*. Signals and Communication Technology. Springer, 2010.

[20] Lorenzo Isella, Juliette Stehlé, Alain Barrat, Ciro Cattuto, Jean-Francois Pinton, and Wouter Van den Broeck. What's in a crowd? analysis of face-to-face behavioral networks. *J. of Theoretical Biology*, 271(1):166–180, 2011.

[21] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabási. The large-scale organization of metabolic networks. *Nature*, 407:651–654, 2000.