Exploring the “Middle Earth” of Network Spectra via a Gaussian Matrix Function

Ernesto Estrada, Alhanouf Ali Alhomaiddhi, Fawzi Al-Thukair

*Department of Mathematics & Statistics, University of Strathclyde, 26 Richmond Street,
Glasgow G11XQ, UK, Department of Mathematics, King Saud University, Saudi Arabia*

We study a Gaussian matrix function of the adjacency matrix of artificial and real-world networks. We motivate the use of this function on the basis of a dynamical process modeled by the time-dependent Schrödinger equation with a squared Hamiltonian. In particular, we study the Gaussian Estrada index—an index characterizing the importance of eigenvalues close to zero. This index accounts for the information contained in the eigenvalues close to zero in the spectra of networks. Such method is a generalization of the so-called “Folded Spectrum Method” used in quantum molecular sciences. Here we obtain bounds for this index in simple graphs, proving that it reaches its maximum for star graphs followed by complete bipartite graphs. We also obtain formulas for the Estrada Gaussian index of Erdős-Rényi random graphs as well as for the Barabási-Albert graphs. We also show that in real-world networks this index is related to the existence of important structural patterns, such as complete bipartite subgraphs (bicliques). Such bicliques appear naturally in many real-world networks as a consequence of the evolutionary processes giving rise to them. In general, the Gaussian matrix function of the adjacency matrix of networks characterizes important structural information not described in previously used matrix functions of graphs.

The spectrum of a network—the set of its eigenvalues—provides important information about the structural and dynamical properties of the corresponding system. Most of the functions used to study network spectra give more weight to the largest modular eigenvalues. Then, the information contained in the eigenvalues close to the centre of the spectra, i.e, those close to zero, has remained totally unexplored in the study of graph spectra. Here we study a Gaussian matrix function that gives more weights to the eigenvalues closest to the centre of the spectrum of a network. It is motivated by the use of the time-dependent Schrödinger equation with a squared Hamiltonian. The time-dependent Schrödinger equation with a normal Hamiltonian is not able to distinguish between the dynamics of a particle hopping among the nodes of a bipartite network from the one hopping on a nonbipartite graph. However, the use of the squared Hamiltonian allows to distinguish them, showing that the wave packet in the bipartite graph is significantly more localized than in the nonbipartite network. Using this Gaussian matrix function we extract important structural information hidden in the spectra of networks, such as the emergence of complete bipartite subgraphs (bicliques), which appear naturally in many real-world networks as a consequence of the evolutionary processes giving rise to them. These bicliques are also ubiquitous in random networks generated by preferential attachment mechanisms, such as the Barabási-Albert model. In this work we provide a series of analytical results that pave the way for further analysis and uses of this Gaussian matrix function to understand network structure and dynamics.

I. INTRODUCTION

Matrix functions [1] have emerged as an important mathematical tool for studying networks [2]. The concepts of communicability [3], subgraph centrality [4, 5] (see also [6] for a review) and Katz index [7] are derived from matrix functions $f(A)$ of the adjacency matrix and allow the characterization of local structural properties of networks. The trace of $f(A)$, which is known as the Estrada index of the graph [8–10], is a useful characterization of the global structure of a graph and it has found applications as an index of natural connectivity for studying robustness of networks [11, 12]. These initial studies have motivated more recent developments in the theory of graph-theoretic matrix function studies [13–15]. All these indices have found multiple applications for studying real-world social, ecological, biological, infrastructural, and technological systems represented by networks [16–18]. Here we will use interchangeably the terms networks and graphs and will follow standard notation as in [17]. The greatest appeal of the use of functions of the adjacency matrix for studying graphs is that when representing them in terms of a Taylor function expansion: $f(A) = \sum_{k=0}^{\infty} c_k A^k$, the entries of the k th power of the adjacency matrix provides information about the number of walks of length k between the corresponding pair of (not necessarily different) nodes (see next section for formal definitions). Then, the important ingredient of the definition of $f(A)$ lies in the use of the coefficients c_k . The use of $c_k = k!^{-1}$ gives rise to the exponential function of the adjacency matrix, which is the basis of the communicability/subgraph centrality. On the other hand, selecting $c_k = \alpha^{-k}$ gives rise to the resolvent of the adjacency matrix, which is the basis of the Katz centrality index [7]. Either of these two coefficients is selected arbitrarily among all the existing possibilities. However, they have proved to be very useful in practice and not very much improvement is obtained by changing the coefficients to account for bigger or smaller penalization of the walks according to their length [19].

Here we propose to investigate the information contained in the mid part of the spectrum of the adjacency matrix of graphs and networks using a new adjacency matrix function. The adjacency matrix of a simple graph always contains positive and negative eigenvalues. Then, we will refer here to the region close to the zero eigenvalue as the middle part of the spectrum. This is only truly the middle part in bipartite networks where the spectrum is symmetric, but we will use the term without loss of generality for any graph. This region of the spectrum is totally unexplored for complex networks. However, there are areas in which the zero eigenvalue plays a fundamental role. For instance, when the adjacency matrix represents the tight-binding Hamiltonian in the Hückel molecular orbital (HMO) method (see [20, 21] for recent reviews), the zero eigenvalue and its multiplicity (graph nullity) represent important parameters related to the molecular stability and molecular magnetic properties (see [22] for a review). In these cases the highest occupied (HOMO) and lowest unoccupied molecular orbitals (LUMO), which correspond to the smallest positive and the smallest negative eigenvalue of A , respectively, play the most fundamental role in the chemical reactivity. It can be said that everything interesting in Chemistry takes place with the involvement of the eigenvalues closest to zero. For instance, many chemical reactions and electron transfer complexes involve electron transfers between the HOMO of one molecule and the LUMO of another [23–25].

Matrix functions of the type of $f(A) = \sum_{k=0}^{\infty} c_k A^k$ are characterized by the fact that they give the highest weight to the largest eigenvalue of the adjacency matrix. For a simple example let us consider the trace of $f(A) = \exp(A)$ of a simple, connected network, which can be written as $\text{Tr} \exp(A) = \sum_{j=1}^n \exp(\lambda_j)$, where n is the order of the graph and $\lambda_1 > \lambda_2 \geq \dots \geq \lambda_n$ are the eigenvalues of A . It is clear that if the spectral gap of the adjacency matrix, $\lambda_1 - \lambda_2$, is very large, $\text{tr} \exp(A)$ depends only on the largest eigenvalue λ_1 . This is not a strange situation in real-world networks, where it is typical to find very large spectral gaps for their adjacency matrix. In these cases the use of functions of the type $f(A)$ makes that the structural information contained in the smaller eigenvalues and eigenvectors of the adjacency matrix is not captured by the index. A similar situation happens if we consider $f(-A)$ [26]. In this case we give more weight to the smallest eigenvalue/eigenvector of the adjacency matrix and the information contained in the largest ones is again lost.

In this work we study a Gaussian adjacency matrix function $f(-A^2)$ as a way to characterize the structural information of graphs giving more importance to the eigenvalues/eigenvectors in the middle part of the graph spectrum. Similar Gaussian operators may arise in quantum mechanics of many body systems [27, 28] as well as the electronic partition function in renormalized tight binding Hamiltonians [29, 30]. We will motivate here the introduction of this function from a quantum-mechanical approach to networks in which a particle is hopping through the nearest neighbor nodes in a graph. We then continue by proving some elementary results for some of the indices derived from $f(-A^2)$ for general graphs. In particular we study here properties of $H = \text{Tr} f(-A^2)$. We show that although the graph nullity—the multiplicity of the zero eigenvalue of the adjacency matrix of the graph—plays an important role in the values of this index, the H index contains more structural information than the graph nullity even for small simple graphs. We then prove that among the graphs with n nodes, the maximum of the H index is always obtained for the star graph followed by other complete bipartite graphs. Then, we obtain analytic expressions for this index in random graphs with Poisson and power-law degree distribution, showing that the last ones always display larger values of the H index than the first ones. Finally, we study more than 60 real-world networks representing

a large variety of complex systems. In this case we study the H index normalized by the network size, $\hat{H}$. We found that the networks with the largest $\hat{H}$ index correspond to those having relatively large bicliques—complete bipartite subgraphs, which can be created by different evolutionary mechanisms depending on the kind of complex system considered. Although there are important network characteristics influencing the $\hat{H}$ index, such as degree distribution and the degree assortativity, we show here that they are not unique in determining the high values of this index observed for certain networks. This new matrix function for graphs and networks may represent an important addition to the characterization of important properties of these systems which have remained unexplored due to the lack of characterizations of the 'middle region' of graph spectra.

II. PRELIMINARIES

Let us introduce some definitions, notations, and properties associated with networks to make this work self-contained. We will use interchangeably the terms graphs and networks in this work. A *graph* $\Gamma = (V, E)$ is defined by a set of n nodes (vertices) V and a set of m edges $E = \{(u, v) | u, v \in V\}$ between the nodes. Here we will consider simple graphs without multiple edges, self-loops and direction of the edges. A *walk* of length k in G is a set of nodes $i_1, i_2, \dots, i_k, i_{k+1}$ such that for all $1 \leq l \leq k$, $(i_l, i_{l+1}) \in E$. A *closed walk* is a walk for which $i_1 = i_{k+1}$. A *path* is a walk with no repeated nodes.

Let A be the adjacency operator on $\ell_2(V)$, namely $(Af)(p) = \sum_{q: \text{dist}(p,q)=1} f(q)$. For simple finite graphs A is the symmetric adjacency matrix of the graph, which has entries

$$a_{uv} = \begin{cases} 1 & \text{if } (u, v) \in E \\ 0 & \text{otherwise} \end{cases} \quad \forall u, v \in V.$$

In the particular case of an undirected network as the ones studied here, the associated adjacency matrix is symmetric, and thus its eigenvalues are real. We label the eigenvalues of A in non-increasing order: $\lambda_1 > \lambda_2 \geq \dots \geq \lambda_n$. Since A is a real-valued, symmetric matrix, we can decompose A into $A = U\Lambda U^T$ where Λ is a diagonal matrix containing the eigenvalues of A and $U = [\vec{\psi}_1, \dots, \vec{\psi}_n]$ is orthonormal, where $\vec{\psi}_i$ is an eigenvector associated with λ_i . Because the graphs considered here are connected, A is irreducible and from the Perron-Frobenius theorem we can deduce that $\lambda_1 > \lambda_2$ and that the leading eigenvector $\mathbf{q}_1$, which will be sometimes referred to as the *Perron vector*, can be chosen such that its components $\vec{\psi}_1(u)$ are positive for all $u \in V$.

Hereafter we will refer to the following function as the communicability function of the graph [2, 3, 6]. Let u and v be two nodes of Γ . The communicability function between these two nodes is defined as

$$G_{uv} = \sum_{k=0}^{\infty} \frac{(A^k)_{uv}}{k!} = (\exp(A))_{uv} = \sum_{k=1}^n e^{\lambda_k} \psi_k(u) \psi_k(v),$$

which is an important quantity for studying communication processes in networks. It counts the total number of walks starting at node u and ending at node v , weighted in decreasing order of their length by a factor $\frac{1}{k!}$; therefore it is considering shorter walks more influential than longer ones. The G_{uu} terms of the communicability function characterize the degree of participation of a node in all subgraphs of the network, giving more weight to the smaller ones. Thus, it is known as the subgraph centrality of the corresponding node [5]. The following quantity is known in the algebraic graph theory literature as the Estrada index of the graph:

$$EE(G) = \sum_{u=1}^n G_{uu} = \text{tr}(\exp(A)) = \sum_{k=1}^n e^{\lambda_k},$$

which is a characterization of the global properties of a network. In its generalized form $EE(G, \beta) = \text{tr}(\exp(\beta A)) = \sum_{k=1}^n e^{\beta \lambda_k}$, it represents the statistical-mechanics partition function of the graph where β represents the inverse temperature.

III. GAUSSIAN ADJACENCY MATRIX FUNCTION OF NETWORKS

A. Motivation

A recent growing area in the study of complex networks is to consider quantum information-theoretic ideas to explore structural properties of networks. As a consequence, the study of entangled networks, centrality, communities and entropies using quantum-mechanical approaches have been developed [31–36]. Here we consider a network $G = (V, E)$ in which there is a particle located at a given node that can hop to another node obeying the rules of quantum mechanics. The quantized energy levels E_j of this system are obtained via the time-independent Schrödinger equation

$$\hat{H}\vec{\psi}_j(\mathbf{r}) = E_j\vec{\psi}_j(\mathbf{r}), \quad (1)$$

where $\hat{H}$ is the Hamiltonian of the system and $\psi_j(\mathbf{r})$ are the wavefunctions. We consider here a simplified version of the Hamiltonian, which is known as the tight-binding model. In this model the Hamiltonian is written as [37]

$$\hat{H}_{TB} = -t \sum c_i^\dagger c_j, \quad (2)$$

where the term $c_i^\dagger c_j$ describes the hopping of a particle from the node i to a nearest neighbor node j via the annihilation of the particle at i and the creation of the particle at j , and t is the hopping integral. Then, it is straightforward to realize that

$$\hat{H} = -A, \quad (3)$$

where A is the adjacency matrix of the network where the particle is hopping and $E_j = -\lambda_j$.

Here we consider an approach known as the *Folded Spectrum Method* (FSM) [38] whose main idea is the following. An eigensolution (E_j, ψ_j) of (1) also satisfies

$$\left(\hat{H} - E_{\text{ref}}\right)^2 \psi_j(\mathbf{r}) = (E_j - E_{\text{ref}})^2 \psi_j(\mathbf{r}), \quad (4)$$

where E_{ref} is a given reference energy. If we select as usual $E_{\text{ref}} = 0$, we have in our tight-binding model

$$A^2 \vec{\psi}_j(\mathbf{r}) = \lambda_j^2 \vec{\psi}_j(\mathbf{r}). \quad (5)$$

This process is ‘folding’ the spectrum of a network is illustrated in Figure (1), which is inspired by a similar one displayed by Canning et al [39]. In the left-hand side of the Figure we shown the normal spectrum of the network, in which the eigenvalues are represented on the line as dots. We then squared the eigenvalues and represent them in a new line on the right-hand side of the Figure. The first consequence of this folding process is that now all the eigenvalues are nonnegative, with the smallest ones being the zeros of the network spectrum. As can be seen in the Figure, the eigenvalues closest to zero, here designated as $\lambda_{\text{ref}\pm 1}$, becomes the lowest state in the spectrum of A^2 . As we will see later this will imply that we can obtain important structural information about the network which is encoded by these eigenvalues of the adjacency matrix.

To give a flavor of the kind of information that A^2 may contain about the network we consider the time-dependent Schrödinger equation based on A^2 ,

$$i \frac{\partial \vec{\psi}(t)}{\partial t} = A^2 \vec{\psi}(t), \quad (6)$$

$$\psi_x(t=0) = \delta_{x,x_0}. \quad (7)$$

The solution of (6) is given by

$$\vec{\psi}(t) = \exp(-itA^2) \vec{\psi}(0). \quad (8)$$

Figure 1. Illustration of the folded spectrum method. The eigenvalues of the adjacency matrix of the network are folded at λ_{ref} into the spectrum of $(A - \lambda_{\text{ref}})^2$. The eigenvalues closest to zero are now the lowest states in the spectrum of $(A - \lambda_{\text{ref}})^2$.

In order to compare the dynamic properties of (6) with that of the time-dependent Schrödinger equation based on A , we will consider the temporal autocorrelation function averaged over different nonequivalent initial positions of the wave packet [40]

$$C(t) = \left\langle \frac{1}{t} \int_0^t |\Upsilon_{pp}(t')|^2 dt' \right\rangle, \quad (9)$$

where

$$\Upsilon_{pq}(t) = [\exp(-itA^2)]_{pq}, \quad (10)$$

and the bracket $\langle \dots \rangle$ indicates the average over the nonequivalent sites. The integrand of this function $P_i(t) = |\Upsilon_{ii}(t)|^2$ is known as the return probability of the wave packet initially localized at the node i of the network. The temporal autocorrelation function describes the return probability of the wave packet to the initial site.

For the time-dependent Schrödinger equation based on A we will have $\Upsilon'_{pq}(t) = [\exp(itA)]_{pq}$ as the solution of that equation is given by $\psi'(t) = \exp(tA)\psi(0)$. In Figure (2) we illustrate the results of the decay of the autocorrelation function in terms of the time t for both Hamiltonians, $\hat{H} = -A$ and $\hat{H} = A^2$ for two cycle graphs with 8 and 9 nodes, respectively. We label all the nodes in consecutive order starting from one. As can be seen in the left panel of Figure (2) there are no significant differences between the decay of the autocorrelation function for both cycles as $t \rightarrow \infty$, in spite of the fact that the graph of C_8 is bipartite and the graph of C_9 is not. We remind the reader that a graph is bipartite if and only if it does not contain any odd cycle. However, when we consider the autocorrelation function based on $\Upsilon_{pq}(t) = [\exp(-itA^2)]_{pq}$ we observe (see right panel of Figure (2)) very significant differences. While the autocorrelation function for C_9 decays to an asymptotic value close to zero as it happens before for $\hat{H} = -A$, the one for the bipartite graph remains with values around 0.4 at $t \rightarrow \infty$. That is, for very long time the probability that the wave packet return to the initial site is very low in the nonbipartite graph, but it is significantly large in the bipartite graph. The difference arises by the fact that in the bipartite graph, a wave packet initially located at node 1 can visit only the nodes 1, 3, 5, 7—similarly a packet initially located at node 2 can only visit 2, 4, 6, 8. That is, the wave packet is significantly more localized in the bipartite graph than in the nonbipartite one, which make it returns more easily to the original site. This property is not easy to observe when using the single adjacency matrix instead of its square.

Figure 2. Plot of the autocorrelation function for the cycles C_8 (dotted line) and C_9 (solid line) using $\Upsilon'_{pq}(t) = [\exp(itA)]_{pq}$ (left panel) and $\Upsilon_{pq}(t) = [\exp(-itA^2)]_{pq}$ (right panel).

We should also notice that the amplitude that a perturbation that resided at the node q at time $t = 0$ ends up at the node p due to the quantum dynamics is given by $\langle p | \exp(-itA^2) | q \rangle$, where we have used Dirac brackets. Then, by continuation from the real time t to the imaginary time, we obtain the thermal Green's function as

$$\tilde{G}_{pq} = (\exp(-\beta A^2))_{pq}. \quad (11)$$

where β is the inverse temperature, hereafter to be considered as $\beta \equiv 1$. The probability that the system, i.e., the network, is found at the state with energy $-\lambda_j^2$ is given by the Boltzmann distribution

$$p_j = \frac{\exp(-\lambda_j^2)}{H}, \quad (12)$$

where the denominator is known as the partition function of the system and it is given by

$$H = \text{Tr}(\exp(-A^2)). \quad (13)$$

In this work we will call

$$\tilde{G} = \sum_{k=0}^{\infty} \frac{(-A^2)^k}{k!} = \exp(-A^2). \quad (14)$$

the *Gaussian matrix function* of A , $\tilde{G}_{pq}$ the *Gaussian communicability function* between the nodes p and q based on $-A^2$ and H the *Gaussian Estrada index* of the graph. The term $\tilde{G}_{pp}$ is the *Gaussian subgraph centrality* based on the same matrix function. In the next section we start the mathematical analysis of these functions.

B. Mathematical Analysis

We start this section by noticing that for calculating the indices previously defined on the basis of the squared-adjacency matrix we do not need to obtain explicitly the exponential matrix of $-A^2$. We are not interested here in the development of such kind of techniques but the reader is directed to the excellent work of Benzi and Boito [41] for

a discussion of efficient techniques for estimating the trace of an exponential matrix that do not require computing every entry of the matrix exponential. Obviously, using the spectral decomposition of the adjacency matrix we can express these indices as

$$\tilde{G}_{pq} = \sum_{j=1}^n \psi_{j,p} \psi_{j,q} \exp(-\lambda_j^2), \quad (15)$$

$$H = \sum_{j=1}^n \exp(-\lambda_j^2). \quad (16)$$

Let $\eta(A)$ be the nullity of the adjacency matrix A , i.e., the dimension of the null space of A . In spectral graph theory $\eta = \eta(A)$ is known as the *graph nullity*. Then, it is obvious that the H index is related to η as follows:

$$H \geq \eta, \quad (17)$$

with both indices identical if and only if $\lambda_j = 0$, for all j , which is attained only for the trivial graph, i.e., the graph with n nodes and no edges. Indeed,

$$H = \eta + \sum_{\lambda_j \neq 0} \exp(-\lambda_j^2). \quad (18)$$

Then, it is interesting at least empirically, to explore the relation between H and η for simple graphs. We investigate all the connected graphs with $n \leq 8$ for which we obtain both H and η . The correlation between both indices for the 11,117 connected graphs with 8 nodes is illustrated in Figure (3). Although the correlation is statistically significant—the Pearson correlation coefficient is 0.74—it hides the important differences between the two indices. For instance, there are 5,724 graphs with zero nullity among all the connected graphs with 8 nodes. For these graphs $1.484 \leq H \leq 3.629$, which represents a wide range of values taking into account that the minimum and maximum values of H for all connected graphs with 8 nodes are 1.484 and 6, respectively. It is also easy to see that there are graphs having nullity zero which have larger H indices than some graphs having nullity one, two or three. The results are very similar for $n < 8$ and they are not shown here. In the Figure (3) we show the graphs with the largest H indices among all connected graphs with 8 nodes and nullity zero or one. These graphs show a common pattern containing several complete bipartite subgraphs. For instance, every yellow node in the Figure (3) is connected to every red ones, every red is connected to every blue and every blue is connected to the green one, while there is no yellow-yellow, red-red or blue-blue connections. This pattern will be revealed when we study the mathematical properties of this index and its importance will be analyzed for real-world networks.

C. General Quadrature Rule-Based Bounds

In this section we will use the Gaussian quadrature rule to obtain an upper bound of H . We will mainly follow here the works [42–44] to which the reader is directed for more. We start by recalling that for a symmetric matrix A with eigendecomposition $A = Q\Lambda Q^T$ and a smooth function f defined on an interval containing the eigenvalues of A , $[a, b]$, for $u, v \in \mathbb{R}^n$ we have:

$$u^T f(A) v = u^T Q f(\Lambda) Q^T v = p^T f(\Lambda) q = \sum_{j=1}^n f(\lambda_j) p_j q_j, \quad (19)$$

where $p = u^T Q$ and $q = Q^T v$. Now we rewrite (19) as a Riemann-Stieltjes integral with respect to the spectral measure:

$$I[f] = u^T f(A) v = \int_b^a f(\lambda) d\mu(\lambda) \text{ where } \mu(\lambda) = \begin{cases} 0, & \lambda \leq a = \lambda_1 \\ \sum_{j=1}^i p_j q_j, & \lambda_i \leq \lambda \leq \lambda_{i+1} \\ \sum_{j=1}^N p_j q_j, & b = \lambda_n \leq \lambda. \end{cases} \quad (20)$$

Figure 3. (a) Plot of graph nullity versus H index for all connected graphs with 8 nodes. (b) Graph with the largest H index among all the connected graphs with 8 nodes and nullity zero. (c) The same as in (b) for all connected graphs with nullity one. Notice in (b) and (c) the connectivity pattern of the graphs in which every yellow rhombus is connected to every red circle, every red circle is connected to every blue square and every blue square is connected to the green triangle one. Also, there is no yellow-yellow (rhombus-rhombus), red-red (circle-circle) or blue-blue (square-square) connections.

is the piecewise constant function.

Our motivation for using this definition is the fact that $[f(A)]_{ij} = e_i^T f(A) e_j$, where e_i is the i th column of the identity matrix. Moreover, $u^T f(A) v = \int_b^a f(\lambda) d\mu(\lambda) = \sum_{j=1}^n w_j f(t_j) + \sum_{k=1}^M v_k f(z_k) + R[f]$ which is the general Gauss-type quadrature rule where the nodes $\{t_j\}_{j=1}^n$ and weights $\{w_j\}_{j=1}^n$ are unknowns, whereas the nodes $\{z_k\}_{k=1}^M$ are prescribed. We have:

- $M = 0$ for the Gauss rule,
- $M = 1$, $z_1 = a$ or $z_1 = b$ for the Gauss-Radau rule,
- $M = 2$, $z_1 = a$ and $z_2 = b$ for the Gauss-Lobatto rule, which we will focus on.

Let J_n be a tridiagonal matrix defined as

$$J_n = \begin{bmatrix} \omega_1 & \gamma_1 & & & \\ \gamma_1 & \omega_2 & \gamma_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \gamma_{n-2} & \omega_{n-1} & \gamma_{n-1} \\ & & & \gamma_{n-1} & \omega_n \end{bmatrix},$$

whose eigenvalues are the Gauss nodes, whereas the Gauss weights are given by the square of the first entries of the normalized eigenvectors of J_n , then,

$$\sum_{l=1}^N w_l f(t_l) = e_1^T f(J_n) e_1. \quad (21)$$

The entries of J_n are computed using the symmetric Lanczos algorithm. Now, if f is a strictly completely monotonic function on an interval $I = [a, b]$ containing the eigenvalues of the matrix A , i.e $f^{(2j)}(x) > 0$ and $f^{(2j+1)}(x) < 0$ on I for all $j \geq 0$ where $f^{(k)}$ denotes the k th derivative of f and $f^{(0)} \equiv f$, the symmetric Lanczos process can be used to compute bounds for the diagonal entries $(f(A))_{ii}$. Let $J_2 = \begin{bmatrix} \omega_1 & \gamma_1 \\ \gamma_1 & \omega_2 \end{bmatrix}$ be the Jacobian matrix obtained by taking a single Lanczos step, then we only need to compute the $(1, 1)$ entry of $f(J_2)$. Now, if $\varphi(x, y) = \frac{\omega_1(f(x) - f(y)) + x f(y) - y f(x)}{x - y}$, then the Gauss-Lobatto rule gives the bound

$$(f(A))_{ii} \leq \varphi(a, b).$$

See [42] for more details.

The main result of this section is the following.

Theorem 1. *Let G be a graph with n nodes and m edges and let $H = \text{tr} \exp(-A^2)$. Then,*

$$H(G) \leq \sum_{i=1}^n \left[\frac{d_i(e^{-b} - 1)}{b} + 1 \right] = 2m \frac{(e^{-b} - 1)}{b} + n. \quad (22)$$

where d_i is the degree of the node i in the graph G , which is the number of edges incident to it, equivalently $d_i = \sum_j a_{ij}$.

Proof. In the case of J_2 we have:

$$J_2 = \begin{bmatrix} \omega_1 & \gamma_1 \\ \gamma_1 & \omega_2 \end{bmatrix}, \omega_1 = a_{ii}, \gamma_1^2 = \sum_{i \neq j} a_{ij}^2, \omega_2 = \frac{1}{\gamma_1^2} \sum_{k \neq i} \sum_{l \neq i} a_{ki} a_{kl} a_{li}.$$

See [42] for more details.

Now, if $B = A^2$, where A is the adjacency matrix of a graph G and $f(x) = e^{-x}$ (which is strictly completely monotonic) we have

$$J_2 = \begin{bmatrix} d_i & \sqrt{\sum_{i \neq j} b_{ij}^2} \\ \sqrt{\sum_{i \neq j} b_{ij}^2} & \frac{1}{\sqrt{\sum_{i \neq j} b_{ij}^2}} \sum_{k \neq i} \sum_{l \neq i} b_{ki} b_{kl} b_{li} \end{bmatrix},$$

where b_{ij} is the (i, j) th entry of A^2 . Notice that $b_{ij} = \sum_{k=1}^n a_{ik} a_{kj}$ and $[a, b] = [0, b]$ since A^2 has nonnegative eigenvalues.

Hence, we have for the Gauss-Lobatto rule

$$(e^{-A^2})_{ii} \leq \frac{d_i(1 - e^{-b}) - b}{-b} = \frac{d_i(e^{-b} - 1)}{b} + 1.$$

To find the bound of the trace of e^{-A^2} we take the summation from 1 to n on the previous inequality, which by the Handshaking Lemma gives the final result. $\square$

IV. H INDEX OF GRAPHS

A. Elementary properties

In the following we show some results about $\tilde{G}_{pq}$ of some elementary graphs which will help us to interpret this measure when applied to more complex structures. In particular, we study the n -nodes path P_n , the n -nodes cycle C_n , the star graph $K_{1,n-1}$, the complete graph K_n of n nodes and the complete bipartite graph K_{n_1,n_2} of $n_1 + n_2$ nodes. P_n is a connected graph in which $n - 2$ nodes are connected to other two nodes and two nodes are connected to only one node; C_n is the connected graph of n nodes in which every node is connected to two others; $K_{1,n-1}$ is the connected graph in which there is one node connected to $n - 1$ nodes, here labeled as 1 and named the central node, and $n - 1$ nodes are connected to the central one only; K_n is the graph in which every pair of nodes is connected by an edge; and K_{n_1,n_2} is the connected graph which is formed by two sets V_1 and V_2 of nodes of cardinalities n_1 and n_2 , respectively, such that every node in V_1 is connected to every node in V_2 . Here we give expressions for the $H(G)$ index of the before mentioned graphs in the form of Lemmas.

Lemma 1. *Let K_n be the complete graph of n nodes. Then*

$$H(K_n) = e^{-(n-1)^2} + \frac{n-1}{e}. \quad (23)$$

Proof. The spectrum of K_n is $\sigma(K_n) = \{[n-1]^1, [-1]^{n-1}\}$ with the eigenvector $\varphi_1 = \frac{1}{\sqrt{n}}(1, 1, \dots, 1)$ so we have

$$\tilde{G}_{pq}(K_n) = \varphi_1(p)\varphi_1(q)e^{-(n-1)^2} + \sum_{j=2}^n \varphi_j(p)\varphi_j(q)e^{-1}, \quad (24)$$

and since the eigenvector matrix has orthonormal rows and columns we have $\sum_{j=2}^n \varphi_j(p)\varphi_j(q) = -\frac{1}{n}$ if $p \neq q$ and $\frac{n-1}{n}$ if $p = q$. Thus,

$$\tilde{G}_{pq}(K_n) = \frac{e^{-(n-1)^2}}{n} - \frac{1}{ne}. \quad (25)$$

Now, if $p = q$ then $\tilde{G}_{pp}(K_n) = \varphi_1^2(p)e^{-(n-1)^2} + \sum_{j=2}^n \varphi_j^2(p)e^{-1} = \frac{e^{-(n-1)^2}}{n} + \frac{n-1}{ne}$.

Then, it is straightforward to realize that

$$H(K_n) = \sum_{j=1}^n \left(\frac{e^{-(n-1)^2}}{n} + \frac{n-1}{ne} \right) \quad (26)$$

$$= e^{-(n-1)^2} + \frac{n-1}{e}. \quad (27)$$

□

Let

$$I_\gamma(x) = \frac{1}{\pi} \int_0^\pi \cos(\gamma\theta) \exp(x \cos \theta) d\theta - \frac{\sin(\gamma\pi)}{\pi} \int_0^\infty \exp(-x \cosh t - \gamma t) dt, \quad (28)$$

be the modified Bessel function of the first kind. Then, we have the following result.

Lemma 2. *Let P_n be a path having n nodes. Then, asymptotically as $n \rightarrow \infty$*

$$H(P_n) = \frac{I_0(2)}{e^2}(n+1) - e^{-4}. \quad (29)$$

Proof. By substituting the eigenvalues and eigenvectors of the path graph into the expression for $\tilde{G}_{pp}(P_n)$ we obtain

$$\tilde{G}_{pp}(P_n) = \frac{2}{n+1} \sum_{j=1}^n \sin^2 \left(\frac{j\pi p}{n+1} \right) \exp \left(-4 \cos^2 \left(\frac{j\pi}{n+1} \right) \right) \quad (30)$$

$$= \frac{e^{-2}}{n+1} \sum_{j=1}^n \left[1 - \cos \left(\frac{2j\pi p}{n+1} \right) \right] \exp \left(-2 \cos \left(\frac{2j\pi}{n+1} \right) \right). \quad (31)$$

Now, when $n \rightarrow \infty$ the summation in 31 can be approached by the following integral

$$\tilde{G}_{pp}(P_n) = \frac{e^{-2}}{\pi} \int_0^\pi \exp(-2 \cos \theta) d\theta - \frac{e^{-2}}{\pi} \int_0^\pi \cos(p\theta) \exp(-2 \cos \theta) d\theta, \quad (32)$$

where $\theta = \frac{2j\pi}{n+1}$. Thus, when $n \rightarrow \infty$ we have

$$\tilde{G}_{pp}(P_n) = e^{-2} (I_0(-2) - I_p(-2)), \quad (33)$$

which by using $I_\gamma(-x) = (-1)^\gamma I_\gamma(x)$ gives

$$\tilde{G}_{pp}(P_n) = e^{-2} (I_0(2) - (-1)^p I_p(2)).$$

Let n be even. Then due to the symmetry of the path we have

$$H(P_n) = 2 \sum_{p=1}^{n/2} \tilde{G}_{pp}(P_n) = 2 \sum_{p=1}^{n/2} e^{-2} [I_0(2) - (-1)^p I_p(2)] \quad (34)$$

$$= \frac{nI_0(2)}{e^2} - \frac{2}{e^2} \sum_{p=1}^{n/2} (-1)^p I_p(2). \quad (35)$$

For $n \rightarrow \infty$ we have

$$\sum_{\gamma=1}^{\infty} (-1)^\gamma I_\gamma(x) = \frac{1}{2} (e^{-x} - I_0(x)). \quad (36)$$

Then, we can write for $n \rightarrow \infty$

$$H(P_n) = \frac{nI_0(2)}{e^2} - \frac{1}{e^2} (e^{-2} - I_0(2)) \quad (37)$$

$$= \frac{I_0(2)}{e^2} (n+1) - e^{-4}. \quad (38)$$

Now, when n is odd we can split the path into two paths of lengths $\frac{n+1}{2}$ and $\frac{n-1}{2}$, respectively. Then, we write

$$H(P_n) = \sum_{p=1}^{\frac{n+1}{2}} \tilde{G}_{pp}(P_n) + \sum_{p=\frac{n-1}{2}}^n \tilde{G}_{pp}(P_n) \quad (39)$$

$$= \frac{(n+1)I_0(2)}{2e^2} - \frac{1}{e^2} \sum_{p=1}^{\frac{n+1}{2}} (-1)^p I_p(2) + \frac{(n-1)I_0(2)}{2e^2} - \frac{1}{e^2} \sum_{p=\frac{n-1}{2}}^n (-1)^p I_p(2). \quad (40)$$

When $n \rightarrow \infty$ we can consider that the summation in the second and fourth terms of 40 are both equal to $(e^{-2} - I_0(2))/2$, which then gives the final result. $\square$

Lemma 3. *Let C_n be a cycle having n nodes. Then, asymptotically as $n \rightarrow \infty$*

$$H(C_n) = \frac{nI_0(-2)}{e^2}. \quad (41)$$

Proof. Notice that the adjacency matrix of a cycle is a circulant matrix and consequently any function of it and that gives

$$H(C_n) = \sum_{j=1}^n \tilde{G}_{jj}, \text{ for any node } j \quad (42)$$

$$= n \left(\frac{\text{tr}(e^{-A^2})}{n} \right) \quad (43)$$

$$= n \left(\frac{1}{n} \sum_{j=1}^n e^{-4 \cos^2(\frac{2\pi j}{n})} \right) \quad (44)$$

$$= ne^{-2} \left(\sum_{j=1}^n \frac{1}{n} e^{-2 \cos \frac{4\pi j}{n}} \right). \quad (45)$$

Now, when $n \rightarrow \infty$ the summation in 45 can be approached by the following integral

$$H(C_n) = ne^{-2} \frac{1}{\pi} \int_0^\pi e^{-2 \cos \theta} d\theta, \quad (46)$$

where $\theta = \frac{2j\pi}{n}$. Thus, when $n \rightarrow \infty$ we have

$$H(C_n) = ne^{-2} I_0(-2). \quad (47)$$

$\square$

Lemma 4. *Let K_{n_1, n_2} be the complete bipartite graph of $n_1 + n_2$ nodes. Then*

$$H(K_{n_1, n_2}) = 2e^{-n_1 n_2} + n_1 + n_2 - 2. \quad (48)$$

Proof. From the orthonormality of the eigenvectors of the adjacency matrix we have:

$$\sum_{j=2}^{n_1+n_2-1} [\varphi_j(p)]^2 = 1 - \frac{1}{n_1}, \quad p \in V_1, \quad (49)$$

$$\sum_{j=2}^{n_1+n_2-1} [\varphi_j(p)]^2 = 1 - \frac{1}{n_2}, \quad p \in V_2. \quad (50)$$

Hence, if $p \in V_1$

$$\tilde{G}_{pp}(K_{n_1, n_2}) = \sum_{j=1}^{n_1+n_2} [\varphi_j(p)]^2 \exp(-\lambda_j^2) \quad (51)$$

$$= e^{-n_1 n_2} \left(\frac{n_1 n_2}{2n_1 n_2^2} + \frac{n_1 n_2}{2n_2 n_1^2} \right) + \sum_{j=2}^{n_1+n_2-1} [\varphi_j(p)]^2 \quad (52)$$

$$= e^{-n_1 n_2} \left(\frac{1}{n_1} \right) + 1 - \frac{1}{n_1} = \frac{1}{n_1} (e^{-n_1 n_2} - 1) + 1, \quad (53)$$

and similarly we have $\tilde{G}_{pp}(K_{n_1, n_2}) = \frac{1}{n_2} (e^{-n_1 n_2} - 1) + 1$ when $p \in V_2$. Then

$$H(K_{n_1, n_2}) = \sum_{j=1}^{n_1+n_2} \tilde{G}_{pp} \quad (54)$$

$$= \sum_{j=1}^{n_1} \tilde{G}_{pp} + \sum_{j=m+1}^{n_1+n_2} \tilde{G}_{pp} \quad (55)$$

$$= n_1 \left(\frac{1}{n_1} (e^{-n_1 n_2} - 1) + 1 \right) + n_2 \left(\frac{1}{n_2} (e^{-n_1 n_2} - 1) + 1 \right) \quad (56)$$

$$= 2e^{-n_1 n_2} + n_1 + n_2 - 2. \quad (57)$$

□

The following corollary will be of importance in the following section of this work.

Corollary 1. *Let $K_{1, n-1}$ be the star graph of n nodes. Then*

$$H(K_{1, n-1}) = 2e^{1-n} + n - 2. \quad (58)$$

B. Graphs with maximum H index

Here we are mainly interested in understanding why certain networks display large values of the H index. Then, we prove that among the graphs with n nodes, the maximum value of the H index is always obtained for the star graph $K_{1, n-1}$. We start this section by proving a general results for trees, which is needed to prove the upper bound.

Lemma 5. *Let T_n be a tree of n nodes, then*

$$H(T_n) \leq H(K_{1, n-1}). \quad (59)$$

Proof. We have the following upper bound

$$H(G) \leq 2m \frac{(e^{-b} - 1)}{b} + n, \quad (60)$$

where m is the number of edges and $[0, b]$ is the interval that contains all the eigenvalues of A^2 . Since A is irreducible then it has a nonnegative real eigenvalue (name it λ_1) which has maximum absolute value among all eigenvalues (Perron-Frobenius).

Now, Let T_n be a tree with $n \geq 2$, then Collatz and Sinogowitz [45] have proved that

$$\lambda_1(T_n) \leq \lambda_1(K_{1,n-1}) = \sqrt{n-1}, \quad (61)$$

where the equality holds if T_n is the star graph. Thus, the interval $[0, n-1]$ contains all the eigenvalues of any tree T_n . Now, substituting in (60)

$$H(T_n) \leq 2(n-1) \frac{(e^{1-n} - 1)}{n-1} + n = n - 2 + 2e^{1-n}. \quad (62)$$

Thus, for any tree of n nodes $H(T_n) \leq H(K_{1,n-1})$. $\square$

Now we prove an important result for general graphs, which also allow us to understand the nature of the index H when studying real-world networks.

Theorem 2. *Let G be connected graph of n nodes, then*

$$H(G) \leq H(K_{1,n-1}). \quad (63)$$

Proof. The largest eigenvalue of any graph G is less than or equal the maximum degree. Thus the interval $[0, (n-1)^2]$ contains all the eigenvalues of A^2 and we get from the quadrature-rule bound

$$H(G) \leq n - 2m \frac{(1 - e^{-(n-1)^2})}{(n-1)^2}. \quad (64)$$

Now, $H(G)$ is maximum when m is the lowest possible for a connected graph. That is,

$$H(G) \leq n - 2 \frac{(1 - e^{-(n-1)^2})}{n-1}. \quad (65)$$

A connected graph with $n-1$ edges is a tree. Then, because of Lemma (5) we have that

$$H(G) \leq H(K_{1,n-1}). \quad (66)$$

$\square$

The previous result is true for any graph with any (finite) n number of nodes. In addition, when $n \rightarrow \infty$, it is easy to see that $H(K_{1,n-1}) \rightarrow n-2$. In a similar way, when $n \rightarrow \infty$

$$H(K_{n_1, n_2}) \rightarrow n_1 + n_2 - 2 = n - 2. \quad (67)$$

Among the graphs with n nodes, as proved here, the maximum value is always reached for the star graph $K_{1,n-1}$. It is then followed by the complete bipartite graph $K_{2,n-2}$, then $K_{3,n-3}$, and so forth. For instance, in the case $n = 8$ we have $H(K_{1,7}) \approx 6.001824$; $H(K_{2,6}) \approx 6.000012$; $H(K_{3,5}) \approx 6.000001$; $H(K_{4,4}) \approx 6.000000$. This observation will play a fundamental role in the analysis of random graphs and real-world networks in the next sections of this work.

Figure 4. Illustration of the graphs having minimum H index among all connected graphs with $n = 4, 5, 6, 7, 8$.

C. Graphs with minimum H index

As we have seen before (see Eq. (18)) the largest contribution to the H index is made by the graph nullity η and by the eigenvalues which are relatively close to zero. Let $x > 0$ be a real number such that $\exp(-x^2) \sim 0$. Then,

$$H \approx \sum_{\substack{\lambda_j \leq x \\ \lambda_j \geq -x}} \exp(-\lambda_j^2). \quad (68)$$

Consequently, the graphs with minimum H index are those having very small density of eigenvalues in the interval $(-x, x)$. For instance, the graph having the smallest H index among all connected graphs with 8 nodes has eigenvalues: -2.0000, -1.7321, -1.0000, -1.0000, -0.8136, 1.4707, 1.7321, 3.3429, which produces $H \approx 1.4845$, which is well approximated if we consider only the eigenvalues in the interval $(-1.5, 1.5)$. The graphs with minimum H index among all connected graphs with $n = 4, 5, 6, 7, 8$ are illustrated in the Figure 4. A complete structural characterization of these graphs is out of the scope of this work, but it calls the attention the existence of bow-tie subgraphs in most of these graphs.

D. H Index of Random Networks

In this section we study two different models of random graphs. They are very ubiquitous as null models for studying real-world networks. The first model is the Erdős-Rényi $G(n, p)$ [46] also known as the Gilbert model [47], in which a graph with n nodes is constructed by connecting nodes randomly in such a way that each edge is included in $G(n, p)$ with probability p independent from every other edge. The second model was introduced by Barabási and Albert [48] on the basis of a preferential attachment process. In this model the graph is constructed from an initial seed of m_0 vertices connected randomly like in an Erdős-Rényi $G(n, p)$. Then, new nodes are added to the network in such a way that each new node is connected to the existing ones with a probability that is proportional to the degree of these existing nodes. While the Erdős-Rényi $G(n, p)$ random graphs have a Poisson degree distribution (when $n \rightarrow \infty$), the Barabási-Albert ones show power-law degree distribution of the form: $p(k) \sim k^{-3}$, where $p(k)$ is the probability of finding a node with degree k . In term of their spectra the main difference is that ER graphs display the Wigner semi-circle distribution [49] of eigenvalues when $n \rightarrow \infty$ of the form

$$\rho(\lambda) = \begin{cases} \frac{2\sqrt{r^2 - \lambda^2}}{\pi r^2} & , -r \leq \lambda \leq r \\ 0, & \text{otherwise,} \end{cases} \quad (69)$$

where $r = 2\sqrt{np(1-p)}$. This normalization factor r has been introduced previously by Farkas et al. [50] for studying the spectra of random networks. However, the BA networks have a triangular distribution [17] of eigenvalues of the form

$$\rho(\lambda) = \begin{cases} \frac{\lambda+r}{r^2}, & -r \leq \lambda < 0 \\ \frac{r-\lambda}{r^2}, & 0 < \lambda \leq r \\ 0 & \text{otherwise.} \end{cases} \quad (70)$$

Using these distributions we obtain the following results.

Theorem 3. *For an Erdős-Rényi random graph $G(n, p)$ with $\frac{\ln n}{n} \ll p$ we have*

$$H(ER) = ne^{-\frac{r^2}{2}} (I_0(\frac{r^2}{2}) + I_1(\frac{r^2}{2})) \quad (71)$$

almost surely, as $n \rightarrow \infty$, where $r = 2\sqrt{np(1-p)}$ and I_n is the modified Bessel function of the first kind.

Proof. We know that the spectral density of $G(n, p)$ converges to the semicircular distribution (69) as $n \rightarrow \infty$. Also, Krivelevich and Sudakov [51] showed that the largest eigenvalue λ_1 of $G(n, p)$ is almost surely $(1 + o(1))np$ provided that $np \gg \ln n$. Then,

$$H(ER) = \exp(-\lambda_1^2) + \sum_{i=2}^n \exp(-\lambda_i^2) \quad (72)$$

$$= e^{-\lambda_1^2} + n \left(\frac{1}{n} \sum_{i=2}^n e^{-\lambda_i^2} \rho(\lambda) \right). \quad (73)$$

When $n \rightarrow \infty$ we have

$$H(ER) = n \int_{-r}^r \rho(\lambda) e^{-\lambda^2} d\lambda \quad (74)$$

$$= \frac{4n}{\pi r^2} \int_0^r \sqrt{r^2 - \lambda^2} e^{-\lambda^2} d\lambda \quad (75)$$

$$= \frac{4n}{\pi r^2} \int_0^{\frac{\pi}{2}} r^2 \cos^2 \theta e^{-r^2 \sin^2 \theta} d\theta \quad (76)$$

$$= \frac{4n}{\pi} \int_0^{\frac{\pi}{2}} \frac{1}{2} (1 + \cos 2\theta) e^{-\frac{r^2}{2}(1 - \cos 2\theta)} d\theta \quad (77)$$

$$= 2ne^{-\frac{r^2}{2}} \left(\frac{1}{\pi} \int_0^{\frac{\pi}{2}} e^{\frac{r^2}{2} \cos 2\theta} d\theta + \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \cos 2\theta e^{\frac{r^2}{2} \cos 2\theta} d\theta \right) \quad (78)$$

$$= ne^{-\frac{r^2}{2}} \left(\frac{1}{\pi} \int_0^\pi e^{\frac{r^2}{2} \cos u} du + \frac{1}{\pi} \int_0^\pi \cos u e^{\frac{r^2}{2} \cos u} du \right) \quad (79)$$

$$= ne^{-\frac{r^2}{2}} (I_0(\frac{r^2}{2}) + I_1(\frac{r^2}{2})). \quad (80)$$

□

We now consider the case of the Barabási-Albert (BA) model as a representative of random graphs with power-law degree distribution. In this case the graph is constructed from an initial seed of n_0 nodes in such a way that new nodes are added in each step. Such new nodes are connected to $m_0 \leq n_0$ existing node. In this case, it is known that $p = \frac{2m_0}{n-1}$. Thus, we can construct the normalization factor $r = 2\sqrt{np(1-p)}$ based on m_0 . We then prove the following result.

Theorem 4. *Let G be a BA random network. Then, when $n \rightarrow \infty$, the H index of a BA network is bounded as*

$$H(BA) = \frac{n}{r^2} \left(\sqrt{\pi} r \operatorname{erf}(r) + e^{-r^2} - 1 \right), \quad (81)$$

where $r = 2\sqrt{np(1-p)}$ and $\operatorname{erf}(\cdots)$ is the error function.

Proof. We know that the density of BA graphs follows a triangular distribution (70). Thus

$$H(BA) = \sum_{j=1}^n \rho(\lambda_j) e^{-\lambda_j^2} \quad (82)$$

$$= n \left(\frac{1}{n} \sum_{j=1}^n \rho(\lambda_j) e^{-\lambda_j^2} \right) \quad (83)$$

$$= n \left(\int_{-r}^r \rho(\lambda) e^{-\lambda^2} d\lambda, \text{ as } n \rightarrow \infty \right) \quad (84)$$

$$= n \left(\int_{-r}^0 \frac{\lambda+r}{r^2} e^{-\lambda^2} d\lambda + \int_0^r \frac{r-\lambda}{r^2} e^{-\lambda^2} d\lambda \right) \quad (85)$$

$$= \frac{n}{r^2} \left(\sqrt{\pi} r \operatorname{erf}(r) + e^{-r^2} - 1 \right). \quad (86)$$

□

In Figure 5(a)) we illustrate the results obtained for the H index of ER random graphs $G_{ER}(1000, p)$ in which p is systematically changed from 0.008 to 0.04. The results are shown for both, the formula (71) and the calculation using the function 'expm' implemented in Matlab®. As can be seen for ER networks, as soon as the probability increases, such that $np \gg \ln n$, the two results quickly converge to a common value, i.e., the error decay quickly with the increase of p . In Figure 5(b)) we also plot similar results for the BA model using $G_{BA}(1000, m_0) = G_{BA}(1000, (n-1)p/2)$ in which m_0 is systematically varied from 2 to 14. In this case we plot the results using the values of p instead of m_0 to make the plot comparable to the one of the ER networks. In this case the behavior is more complex as there is a crossing point between the two curves. This difference between the behavior of the theoretical function (81) for low and large densities of the graphs may be due to the fact that the eigenvalue distribution of the BA networks is different at these two density regimes. According to our computational experiments, it is only true that the BA networks display triangular eigenvalue distributions for relatively small edge densities and deformations of it occurs for larger densities, which may produce the observed deviations from the theoretical and computational results. More theoretical work is needed to understand completely the eigenvalue distribution of these networks at different density regimes. Such studies are clearly out of the scope of the current work.

It is easy to show that for a given value of r , $H(BA) > H(ER)$. That is, for the same network density the network having power-law degree distribution has larger value of the H index than the analogous one with Poisson degree distribution. This result is somehow expected from the qualitative analysis of the eigenvalues distributions of these two classes of random networks. While the ER networks display a semicircle distribution of eigenvalues, the BA networks for small values of r displays a triangular distribution peaked at $\lambda_j = 0$. In other words, the nullity of the BA graphs is larger than that of the ER ones, and the concentration of eigenvalues close to zero is also larger for the BA networks than for the ER. Both characteristics give rise to larger values of the H index in the BA networks. The question that arises here is what this difference implies from the structural point of view. We will analyze this question in the remaining part of this section.

We have already seen that the largest values of the H index occurs in graphs having complete bipartite structures. Then, in order to understand the main structural differences giving rise to the larger H index in BA networks than in ER ones we consider the existence of such subgraphs in both networks. In particular, we will consider the existence of complete bipartite subgraphs, known as bicliques, in both kind of networks. In the current work we will give only a qualitative explanation of this difference which will point to the direction of a further quantitative analysis. Let us start by the analysis of the BA networks. These networks are created from an initial seed of n_0 nodes connected randomly and independently according to the ER model. Then, at each stage of the evolution of the network, a new node is connected preferentially to $m_0 \leq n_0$ nodes. The connection probability is proportional to the degree of the existing nodes. Because an ER network is uncorrelated the probability that the highest degree nodes are connected to each other is relatively low. Then, when a new node is added and connected to m_0 of the highest degree existing nodes there is a high probability that a biclique is formed. Such a process is continued as more nodes are added to the graph, resulting a large bicliques with high probability (see Figure 6). The creation of an ER network follows a completely different process in which pairs of nodes are connected randomly and independently, which does not generates any preferred subgraphs, thus not producing a large number of bicliques. This qualitative analysis explaining structurally the existence of networks with high values of the H index will be very useful in the next section of this work where we will analyze real-world networks.

Figure 5. (a) Change of the H index with the increase of the probability p in ER random graphs $G_{ER}(1000, p)$ obtained using the formula (71) (empty circles and solid line) and using the function 'expm' in Matlab (squares and broken line). (b) Change of the H index with the increase of m_0 in BA random graphs $G_{BA}(1000, m_0)$ obtained using the formula (81) (empty circles and solid line) and using the function 'expm' in Matlab (squares and broken line). All the calculations are the average of 100 random realizations.

Figure 6. Illustration of the evolution of a graph under the BA model to sketch how bicliques are formed in such kind of networks. (a) Seed of $n_0 = 7$ nodes created with a Poissonian degree distribution to start the BA evolution process. (b) Given $m_0 = 2$ the new node (red rhombus one) is preferentially attached to those with the highest degree among the existing n_0 ones (marked as blue squares). (c) Second iteration of the process, which creates a biclique $K_{2,2}$ (red rhombus and blue squares nodes joined by dotted lines).

In order to corroborate the hypothesis that the presence of complete bipartite structures on networks play a fundamental role on the high values of the H index we developed the following experiment. We consider complete bipartite graphs K_{n_1, n_2} with $n = n_1 + n_2$ equal to 100, and 200, respectively. Then, we rewired randomly and independently a small percentage of the edges of these graphs and compute the average value of the H index after 100 random rewirings. These rewirings are produced by keeping the same degree of each of the nodes in the original graphs. In Figure (7) we illustrate graphically the results of this experiment. We keep the rewiring probability low to avoid the appearance of other effects produced by the total neglection of the original complete bipartite structure. As can be seen for each of the complete bipartite graphs studied, the rewiring of the edges, which necessarily implies the destruction of the complete bipartivity, decays significantly the values of H index. We have studied two sizes of the graphs and observed that there are no significant differences in the results when the size of the graphs is doubled from 100 to 200. However, there is a significant change in the shape of the decaying function of the H index as a function of the rewiring probability for different kinds of complete bipartite structures. That is, when $n_1 = n_2$ there

Figure 7. Effects of rewiring the edges of complete bipartite graphs K_{n_1, n_2} by keeping the degree of the nodes on the H index. In the left panel we illustrate graphs with 100 nodes and in the right panel we shown results for graphs with 200 nodes. Every point in the curves is obtained from the average of 100 random rewirings of the edges of each graph that keeps the same degree of the nodes.

is a much faster decay to the H index as a consequence of the rewiring than when $n_1 \gg n_2$. In the last case, the number of nodes in one of the two disjoint sets of the bipartite graphs is relatively small in relation to the number of edges rewired. Consequently, these nodes can be totally saturated by a few rewirings that connect almost every pair of the nodes in this set. Thus, such small set is transformed into a random subgraph more quickly than a larger set of nodes, which is the case when $n_1 = n_2$.

In closing, we have provided theoretical and empirical evidence showing that the H index accounts for the existence of complete bipartite subgraphs in a network. Such subgraphs can appear naturally in preferential attachment processes, like the Barabási-Albert model, but they may also emerge as a consequence of other natural or man-made processes giving rise to the actual structure of networks. We will explore these possibilities in the next section of this paper.

V. STUDIES OF REAL-WORLD NETWORKS

A. Datasets

In this section we study a group of real-world networks representing a variety of social, environmental, technological, infrastructural and biological complex systems. A description of the networks and their main characteristics are given below.

Brain networks

- Neurons: Neuronal synaptic network of the nematode *C. elegans*. Included all data except muscle cells and using all synaptic connections [52]; Cat and macaque visual cortices: the brain networks of macaque visual cortex and cat cortex, after the modifications introduced by Sporn and Kötter [53].

Ecological networks

- Benguela: Marine ecosystem of Benguela off the southwest coast of South Africa [54]; Bridge Brook: Pelagic species from the largest of a set of 50 New York Adirondack lake food webs [55]; Canton Creek: Primarily invertebrates and algae in a tributary, surrounded by pasture, of the Taieri River in the South Island of New Zealand [56]; Chesapeake Bay: The pelagic portion of an eastern U.S. estuary, with an emphasis on larger

fishes [57]; Coachella: Wide range of highly aggregated taxa from the Coachella Valley desert in southern California [58]; El Verde: Insects, spiders, birds, reptiles and amphibians in a rainforest in Puerto Rico [59]; Grassland: all vascular plants and all insects and trophic interactions found inside stems of plants collected from 24 sites distributed within England and Wales [60]; Little Rock: Pelagic and benthic species, particularly fishes, zooplankton, macroinvertebrates, and algae of the Little Rock Lake, Wisconsin, U.S. [61]; Reef Small: Caribbean coral reef ecosystem from the Puerto Rico-Virgin Island shelf complex [62]; Scotch Broom: Trophic interactions between the herbivores, parasitoids, predators and pathogens associated with broom, *Cytisus scoparius*, collected in Silwood Park, Berkshire, England, UK [63]; Shelf: Marine ecosystem on the northeast US shelf [64]; Skipwith: Invertebrates in an English pond [65]; St. Marks: Mostly macroinvertebrates, fishes, and birds associated with an estuarine seagrass community, *Halodule wrightii*, at St. Marks Refuge in Florida [66]; St. Martin: Birds and predators and arthropod prey of *Anolis* lizards on the island of St. Martin, which is located in the northern Lesser Antilles [67]; Stony Stream: Primarily invertebrates and algae in a tributary, surrounded by pasture, of the Taieri River in the South Island of New Zealand in native tussock habitat [68]; Ythan_1: Mostly birds, fishes, invertebrates, and metazoan parasites in a Scottish Estuary [69]; Ythan_2: Reduced version of Ythan1 with no parasites [70].

- Termite: The networks of three-dimensional galleries in termite nests [71]; Ant: The network of galleries created by ants [72]; Dolphins: social network of frequent association between 62 bottlenose dolphins living in the waters off New Zealand [73];

Informational networks

- Centrality: Citation network of papers published in the field of Network Centrality [74, 75]; GD: Citation network of papers published in the Proceedings of Graph Drawing during the period 1994-2000 [76]; ODLIS: Vocabulary network of words related by their definitions in the Online Dictionary of Library and Information Science. Two words are connected if one is used in the definition of the other [77]; Roget: Vocabulary network of words related by their definitions in Roget's Thesaurus of English. Two words are connected if one is used in the definition of the other [78]; Small World: Citation network of papers that cite S. Milgram's 1967 Psychology Today paper or use Small World in title [79].

Biological networks

- Protein-protein interaction networks in: *Kaposi sarcoma herpes virus* (KSHV) [80]; *P. falciparum* (malaria parasite) [81]; *S. cerevisiae* (yeast) [82, 83]; *A. fulgidus* [84]; *H. pylori* [85]; *E. coli* [86] and *B. subtilis* [87].
- Trans_E.coli: Direct transcriptional regulation between operons in *Escherichia coli* [88, 89]; Trans_sea_urchin: Developmental transcription network for sea urchin endomesoderm development. [88]; Trans_yeast: Direct transcriptional regulation between genes in *Saccaromyces cerevisiae*. [52, 88].

Social and economic networks

- Corporate: American corporate elite formed by the directors of the 625 largest corporations that reported the compositions of their boards selected from the Fortune 1000 in 1999 [90]; Geom: Collaboration network of scientists in the field of Computational Geometry [79]; Prison: Social network of inmates in prison who chose "What fellows on the tier are you closest friends with?" [91]; Drugs: Social network of injecting drug users (IDUs) that have shared a needle in the last six months [92]; Zachary: Social network of friendship between members of the Zachary karate club [93]; College: Social network among college students in a course about leadership. The students choose which three members they wanted to have in a committee [94]; ColoSpring: The risk network of persons with HIV infection during its early epidemic phase in Colorado Spring, USA, using analysis of community wide HIV/AIDS contact tracing records (sexual and injecting drugs partners) from 1985-1999 [95]; Galesburg: Friendship ties among 31 physicians [75]; High_Tech: Friendship ties among the employees in a small high-tech computer firm which sells, installs, and maintain computer systems [75, 96]; Saw Mills: Social communication network within a sawmill, where employees were asked to indicate the frequency with which they discussed work matters with each of their colleagues [75, 97];

Technological and infrastructural networks

- Electronic: Three electronic sequential logic circuits parsed from the ISCAS89 benchmark set, where nodes represent logic gates and flip-flop [52]; USAir97: Airport transportation network between airports in US in 1997 [79]; Internet: The internet at the Autonomous System (AS) level as of September 1997 and of April 1998 [98]; Power Grid: The power grid network of the Western USA [99].

n	H	EE	η	r
	0.942	0.283	0.734	0.203
		0.202	0.915	0.064
			0.030	-0.027
				-0.077

Table I. Pearson correlation coefficients among every pair of variables studied in this work (see (V B)).

Software networks

- Collaboration networks associated with six different open-source software systems, which include collaboration graphs for three Object Oriented systems written in C++, and call graphs for three procedural systems written in C. The class collaboration graphs are from version 4.0 of the VTK visualization library; the CVS snapshot dated 4/3/2002 of Digital Material (DM), a library for atomistic simulation of materials; and version 1.0.2 of the AbiWord word processing program. The call graphs are from version 3.23.32 of the MySQL relational database system, and version 1.2.7 of the XMMS multimedia system. Details of the construction and/or origin of these networks are provided in Myers [100].

B. Analysis of real-world networks

The sizes of the networks studied here range from 29 to 4,941 nodes. Together with the H index we have also calculated the Estrada index $EE = tr(\exp(A))$, the graph nullity η , and the Newman degree assortativity coefficient r [101]. We then obtained the linear correlation coefficient among every pair of variables for the group of networks studied and the results are given in Table (V B).

As can be seen the H index displays a significant correlation with the network size. Then, in order to avoid any size influence, we normalize the H index by dividing it by the number of nodes of the network. We will call $\hat{H}$ to the normalized index. The normalized index $\hat{H}$ ranges from about 0.14 to about 0.75 for the studied networks, indicating that real-world networks cover most of the values that this index can take (see V B). The scatterplot of the normalized nullity versus the normalized H index for the 61 real-world networks studied here (plot not shown) reveals that although both indices follow the same trend, there are important differences among them. In particular, we can observe that there are 9 networks with zero nullity which display values of $\hat{H}$ ranging from about 0.14 (the lowest $\hat{H}$ index) to about 0.36 (ranked 25th in increasing order of $\hat{H}$ index).

Name	H	$\hat{H}$	EE	η	r
Ants	30.998	0.4189	2.64E+02	14	-0.102
Benguela	9.573	0.3301	4.11E+06	0	0.021
BridgeBrook	56.018	0.7469	9.20E+08	48	-0.668
Canton	40.333	0.3735	3.12E+08	24	-0.226
CatCortex	12.636	0.2430	8.95E+09	0	-0.044
Centrality_literature	42.976	0.3642	2.44E+08	9	-0.202
Chesapeake	13.240	0.4012	4.71E+02	3	-0.196
Coachella	10.984	0.3661	7.61E+07	0	0.035
ColoSpg	182.077	0.5620	1.15E+03	142	-0.295
CorporatePeople	228.395	0.1440	1.27E+10	0	0.268
Dolphins	20.845	0.3362	2.06E+03	2	-0.044
Drugs	279.467	0.4537	6.91E+07	131	-0.117
Electronic1	37.694	0.3090	4.84E+02	0	-0.002
Electronic2	77.982	0.3095	1.04E+03	8	-0.006
Electronic3	158.658	0.3099	2.17E+03	24	-0.030
ElVerde	51.696	0.3314	4.76E+13	5	-0.174
Galesburg	9.519	0.3071	4.36E+02	1	-0.135
GD	90.440	0.3632	1.60E+04	15	0.098
Geom	1462.396	0.4039	4.04E+12	537	0.168
Hi_tech	10.975	0.3326	2.95E+03	1	-0.087
Internet1997	2148.635	0.7126	6.17E+13	1883	-0.229
Internet1998	2473.122	0.7022	1.42E+15	2158	-0.210
LittleRockA	117.772	0.6507	5.32E+17	93	-0.234
MacaqueVisualCortex	9.665	0.3020	1.26E+06	1	0.008
Neurons	69.083	0.2467	1.31E+10	3	-0.069

ODLIS	1131.046	0.3903	1.54E+19	270	-0.173
PIN_Afulgidus	16.366	0.5114	9.91E+01	12	-0.472
PIN_Bsubtilis	53.144	0.6327	3.52E+02	46	-0.486
PIN_Ecoli	102.189	0.4443	8.30E+06	57	-0.015
PIN_Hpyroli	397.649	0.5601	4.60E+04	316	-0.243
PIN_KSHV	18.119	0.3624	1.82E+03	2	-0.058
PIN_Malaria	83.377	0.3641	2.25E+04	13	-0.083
PIN_Yeast	1135.731	0.5107	1.94E+08	754	-0.105
Power_grid	1907.307	0.3860	2.13E+04	593	0.003
PRISON	20.325	0.3034	7.08E+02	0	0.103
ReefSmall	12.888	0.2578	2.07E+10	0	-0.193
Roget	264.570	0.2662	2.38E+05	2	0.174
Sawmill	12.307	0.3419	2.57E+02	2	-0.071
ScotchBroom	103.975	0.6752	2.46E+06	90	-0.311
Shelf	20.724	0.2559	1.60E+18	2	-0.094
Skipwith	15.023	0.4292	3.87E+09	7	-0.319
SmallWorld	115.730	0.4967	1.27E+09	70	-0.303
College	8.049	0.2515	5.36E+02	0	-0.119
Software_Abi	575.133	0.5557	1.65E+05	418	-0.086
Software_Digital	82.277	0.5485	1.31E+03	63	-0.228
Software_Mysql	648.971	0.4385	2.70E+09	282	-0.083
Software_VTK	440.251	0.5710	1.11E+05	324	-0.195
Software_XMMS	478.168	0.4924	4.64E+04	294	-0.114
StMarks	13.607	0.2835	1.43E+05	0	0.111
StMartin	14.438	0.3281	2.78E+05	2	-0.153
Stony	41.359	0.3693	7.23E+09	30	-0.222
Termite_1	206.581	0.4075	1.92E+03	75	-0.046
Termite_2	116.912	0.4497	7.32E+02	58	-0.150
Termite_3	100.975	0.3768	1.89E+03	23	0.045
Trans_Ecoli	214.517	0.6540	1.06E+04	184	-0.265
Trans_urchin	22.218	0.4937	9.12E+02	13	-0.207
Transc_yeast	478.315	0.7225	3.59E+04	440	-0.410
USAir97	142.765	0.4300	8.08E+17	58	-0.208
Ythan1	58.374	0.4356	1.86E+07	23	-0.263
Ythan2	41.326	0.4492	7.07E+06	22	-0.322
Zackar	15.994	0.4704	1.04E+03	10	-0.476

Table 2. Dataset of real-world networks studied in this paper, their size n , Gaussian Estrada index H , exponential Estrada index EE , graph nullity η , and degree assortativity r .

The largest value of $\hat{H}$ corresponds to the food web of Bridge Brook, which displays the second highest normalized nullity, i.e., the nullity divided by n . It is followed by the transcription network of yeast (displaying the highest value of the normalized nullity) and the versions of Internet at Autonomous System (AS) of 1997 and 1998. The three networks display triangular eigenvalue distributions peaked at the zero eigenvalue which explains their large values of the $\hat{H}$ index. However, while the yeast transcription network and the Internet at AS have fat-tailed degree distributions, the Bridge Brook food web displays a uniform one. Thus, the existence of large values of the $\hat{H}$ index is not tied up to the existence of fat-tailed degree distributions. Most of the networks (75.4%) have values of the $\hat{H}$ index below 0.5. That is, only 15 networks out of 61 have $\hat{H} \geq 0.5$. Among these 15 networks there are 4 of the 7 protein-protein interaction networks (PINs) studied and two of the three transcription networks studied. Thus, almost half of the networks with $\hat{H} \geq 0.5$ represent biological systems containing proteomic or transcriptomic information. The other transcription network studied has $\hat{H} \approx 0.494$ and the other 3 PINs have values of $\hat{H}$ ranging between 0.36 and 0.44. It is interesting to explore the main structural causes for these high values of the $\hat{H}$ index. In previous sections we have found that one of the main structural characteristics determining the high values of this index is the presence of bicliques, e.g. the highest value of $\hat{H}$ is obtained for complete bipartite graphs, also the BA networks display larger $\hat{H}$ index than the ER ones due to the presence of complete bipartite subgraphs created during the evolution of the preferential attachment mechanism. Consequently, we should expect that such kind of subgraphs appear in those real-world networks having the largest $\hat{H}$ index. In the case of the food web of Bridge Brook we have found a biclique consisting of two sets of nodes V_1 and V_2 with cardinalities of 6 and 35 nodes, respectively (see Figure 8 top panel). This subgraph represents a biclique $K_{6,35}$ which contains 55% of the total number of nodes in the network. There are also other smaller bicliques in this network (one of them is a biclique $K_{7,15}$), which together with the $K_{6,35}$ contribute to the large $\hat{H}$ value observed. In Figure (8 (bottom left panel)) we illustrate the decay of the temporal

autocorrelation function (9) based on the Schrödinger equation (4) with the increase of time for the network of Bridge Brook (continuous line) and the totally rewired version of it that preserves the degree of the nodes of the original network (dotted line). As can be seen, the wave packet hopping among the nodes of the real-world network returns to the initial site with a higher probability than in the totally rewired network. That is, in the original network, the hopping particle get stacked in certain regions of the network due to the presence of large complete bipartite subgraphs. Such localization of the wave packet is not observed at all in the totally rewired network. In the bottom right panel of Figure (8) we plot similar results for the network of the cat visual cortex, which has a low value of the H index. In this case we can see that the real network and the totally random version of it are not very much different from each other in their quantum transport using the Schrödinger equation (4). We have also corroborated the differences between these two kind of systems by plotting the decay of the H index as a function of the rewiring probability (plots not shown). In this case we observed that the network of Bridge Brook displays a significant decay of the H index for relatively small rewiring probabilities, while the H index of the cat visual cortex network remains almost constant after random rewiring.

In the cases of the yeast transcription network and the Internet at AS, the networks are characterized by having a few hubs connected to many nodes of degree one, then producing bicliques of the type K_{1,n_2} . In general these findings can be understood on the basis of different mechanisms which give rise to the existence of bicliques in real-world networks. For instance, in some food webs there are top predators which compete for a group of preys. If for this group of species there are no prey-prey nor predator-predator trophic interactions, the corresponding subgraph is a biclique as the one observed for the Bridge Brook network previously considered. In the cases of transcription and PINs the bicliques can be formed as a consequence of lock-and-key kind of interaction. That is, a group of proteins (genes) can act as locks (activators) that physically interact with other proteins (activate other genes) acting as keys. Such kind of interactions is prone to produce relatively large bicliques in the structure of the networks resulting from them. On the other hand, among the networks with $\hat{H} \leq 0.3$ we find the network of corporate directors, the three neuronal networks studied, i.e., macaque and cat visual cortex and the neuronal network of *C. elegans*, as well as some social networks and food webs. Also, the three electronic circuits studied here also display values of $\hat{H}$ index around 0.3. These networks are characterized by the lack of complete bipartite subgraphs and they may represent a variety of topologies difficult to be reproduced by a single mechanism.

Finally we would like to remark a few important characteristics of the Gaussian matrix function of a network that point out to the necessity of further studies of it for real-world networks and simple graphs in general. The first, is our observation that although networks with fat-tailed degree distribution may give rise to high values of the $\hat{H}$ index, it is not a necessary condition for a network to display such a characteristic. We have seen that networks with exponential and even uniform degree distributions display large values of the $\hat{H}$ index. Another structural parameter that could be related to the $\hat{H}$ index is the degree assortativity, i.e., the Pearson correlation coefficient of the degree-degree distribution of a network. We have explored such relation between the $\hat{H}$ index and the assortativity for the 61 networks studied here. We have found that the two parameters are negatively correlated. That is, high values of the $\hat{H}$ index in general implies that the networks are disassortative, i.e., there is a trend of high degree nodes to be connected to low degree ones. This is understandable on the basis of our findings that bicliques of the type K_{1,n_2} plays a fundamental role in the value of the $\hat{H}$ index. However, the correlation is very weak and displays a Pearson correlation coefficient of -0.68. Thus, further explorations—both theoretical and computational—of the relation of the $\hat{H}$ index and other network parameters are necessary for a complete understanding of this index and its application in network theory.

VI. CONCLUSIONS

Most of the works using matrix functions for studying graphs are concentrated on the use of the exponential and the resolvent of the adjacency matrix of the graph. Other functions such as the hyperbolic sine and cosine, and ψ -matrix functions have also been reported. All these matrix functions give more weight to the largest eigenvalue and the corresponding eigenvector of the adjacency matrix than to the rest of eigenvalues/eigenvectors. In many real-world networks, where the spectral gap is relatively large, this situation gives rise to discarding important structural information contained in the eigenvalues close to zero in the graph spectra. Here, we have studied a Gaussian matrix function which accounts for the information contained in the eigenvalues/eigenvectors close to zero in the graph spectra. We have shown that such information is related to the existence of important structural patterns in graphs which have remained unexplored when studying the structure of complex networks, such as the existence of relatively large complete bipartite subgraphs (bicliques). Such bicliques appear naturally in many real-world networks as well as in the Barabási-Albert graphs and other networks with fat-tailed degree distributions. In this work we have concentrated in the theoretical characterization of the networks displaying the largest Gaussian Estrada index—an index characterizing the importance of eigenvalues close to zero. Other extensions to give more weight to other specific

Figure 8. (top) Representation of the food web of Bridge Brook. The 35 nodes in layer 1 and the six nodes in layer 2 form a biclique $K_{6,35}$. Layers 3 and 4 form a biclique $K_{7,15}$. (bottom, left panel) Change in the temporal autocorrelation function based on $\Upsilon_{pq}(t) = [\exp(-itA^2)]_{pq}$ for the food web of Bridge Brook (solid blue line) and in a totally rewired version of that network that preserves the node degree (red broken line). (bottom right panel) The same as before, for the brain network of the cat visual cortex.

eigenvalues/eigenvectors of the adjacency matrix are under development. We hope this work will open new research interest in the study of matrix functions for the structural characterization of graphs.

ACKNOWLEDGMENTS

EE thanks the Royal Society of London for a Wolfson Research Merit Award. The authors thank Dr Manlio Di Domenico (Tarragona) for the illustration of the network in Figure 8.

-
- [1] N. J. Higham, *Functions of Matrices: Theory and Computation*. (Society for Industrial and Applied Mathematics, Philadelphia, PA, 2008).
 - [2] E. Estrada, and D. J. Higham, "Network properties revealed through matrix functions," *SIAM Rev.* **52**, 696–714 (2010).
 - [3] E. Estrada, and N. Hatano, "Communicability in complex networks," *Phys. Rev. E* **77**, 036111 (2008).
 - [4] E. Estrada, "Characterization of the folding degree of proteins," *Bioinformatics* **18**, 697–704 (2002).
 - [5] E. Estrada, and J.A. Rodríguez-Velázquez, "Subgraph centrality in complex networks," *Phys. Rev. E*, **71**, 056103 (2005).
 - [6] E. Estrada, N. Hatano, and M. Benzi, "The physics of communicability in complex networks," *Phys. Rep.* **514**, 89–119 (2012).
 - [7] L. Katz, "A new index derived from sociometric data analysis," *Psychometrika* **18**, 39–43 (1953).
 - [8] J.A. de la Peña, I. Gutman, and J. Rada, "Estimating the Estrada index," *Lin. Algebra Appl.* **427**, 70–76 (2007).
 - [9] H. Deng, S. Radenković, and I. Gutman, "The Estrada index," in *Applications of Graph Spectra*, edited by D. Cvetković, I. Gutman, (Math. Inst, Belgrade, 2009), pp. 123–140.
 - [10] I. Gutman, H. Deng, and S. Radenković, "The Estrada index: an updated survey," in *Selected Topics on Applications of Graph Spectra*, edited by D. Cvetković, I. Gutman, (Math. Inst, Belgrade, 2011), pp. 155–174.
 - [11] J. Wu, M. Barahona, Y.-J. Tan, and H.-Z. Deng. "Robustness of regular ring lattices based on natural connectivity," *Int. J. Syst. Sci.* **42**, 1085–1092, 2011.
 - [12] J. Wu, H.-Z. Deng, Y.-J. Tan, and D.-Z. Zhu, "Vulnerability of complex networks under intentional attack with incomplete information," *J. Phys. A: Math. Theor.* **40**, 2665 (2007).
 - [13] M. Benzi, E. Estrada, and C. Klymko, "Ranking hubs and authorities using matrix functions," *Lin. Algebra Appl.*, **438**, 2447–2474(2013).
 - [14] M. Benzi, and C. Klymko, "On the limiting behavior of parameter-dependent network centrality measures," *SIAM J. Matrix Anal. Appl.* **36**, 686–706 (2015).
 - [15] F. Arrigo, and M. Benzi, "Updating and downdating techniques for optimizing network communicability," *SIAM J. Sci. Comp.* **38**, B25–B49 (2016).
 - [16] E. Estrada, "Graphs and Networks," in *Mathematical Tools for Physicists*, edited by M. Grinfeld, (John Wiley & Sons, 2014), pp. 111–157
 - [17] E. Estrada, *The Structure of Complex Networks. Theory and Applications*. (Oxford University Press, 2011).
 - [18] L. F. Costa, O. N. Oliveira Jr, and G. Travieso, *et al.* "Analyzing and modeling real-world phenomena with complex networks: a survey of applications," *Adv. Phys.* **60**, 329–412 (2011).
 - [19] E. Estrada, "Generalized walks-based centrality measures for complex biological networks," *J. Theor. Biol.* **263**, 556–565 (2010).
 - [20] W. Kutzelnigg, "What I like about Hückel theory," *J. Comput. Chem.* **28**, 25–34 (2007).
 - [21] K. Yates, *Hückel molecular orbital theory*. (Elsevier, 2012).
 - [22] B. Borovičanin, and I. Gutman, "Graph nullity," in *Selected Topics on Applications of Graph Spectra*, edited by D. Cvetković, I. Gutman, (Math. Inst, Belgrade, 2011), pp. 155–174.
 - [23] K. Fukui, T. Yonezawa, H. Shingu, "A Molecular Orbital Theory of Reactivity in Aromatic Hydrocarbons," *J. Chem. Phys.* **20**, 722–725 (1952).
 - [24] K. Fukui, "The role of frontier orbitals in chemical reactions (Nobel Lecture)," *Angewandte Chemie Int. Ed. Engl.* **21**, 801–809 (1982).
 - [25] I. Fleming, *Frontier Orbitals and Organic Chemical Reactions*. (London: Wiley, 1978), pp. 29–109.
 - [26] E. Estrada, D. J. Higham, and N. Hatano, "Communicability betweenness in complex networks," *Physica A* **388**, 764–774 (2009)
 - [27] J. E. Hirsch, and J. R. Schrieffer, "Dynamic correlation functions in quantum systems: A Monte Carlo algorithm," *Phys. Rev. B* **28**, 5353–5356 (1983).
 - [28] E. R. Gagliano and C. A. Balseiro, "Dynamic correlation functions in quantum many-body systems at zero temperature," *Phys. Rev. B* **38**, 11766–11773 (1988).
 - [29] J. E. Barrios-Vargas and G. G. Naumis, "Doped graphene: the interplay between localization and frustration due to the underlying triangular symmetry," *J. Phys.: Condens. Matter* **23**, 375501 (2011).
 - [30] E. Estrada, and M. Benzi, "Atomic displacements due to spin-spin repulsion in conjugated alternant hydrocarbons," *Chem. Phys. Lett.* **568–569**, 184–189 (2013).
 - [31] A. Acn, J. I. Cirac, and M. Lewenstein, "Entanglement percolation in quantum networks," *Nature Phys.* **3**, 256–259 (2007).

- [32] M. Cuquet and J. Calsamiglia, “Entanglement percolation in quantum complex networks,” *Phys. Rev. Lett.* **103**, 240503 (2009).
- [33] S. Perseguers, M. Lewenstein, A. Acn, and J. I. Cirac, “Quantum random networks,” *Nature Phys.* **6**, 539-543 (2010).
- [34] M. Faccin, T. Johnson, J. Biamonte, S. Kais, and P. Migdal, “Degree distribution in quantum walks on complex networks,” *Phys. Rev. X* **3**, 041007 (2013).
- [35] M. Faccin, P. Migdal, T. H. Johnson, V. Bergholm, and J. D. Biamonte, “Community detection in quantum complex networks,” *Phys. Rev. X* **4**, 041012 (2014).
- [36] M. De Domenico and J. Biamonte, “Spectral entropies as information-theoretic tools for complex network comparison,” *arXiv:1609.01214* (2016).
- [37] E. Canadell, M.-L. Doublet, Ch. Iung, *Orbital Approach to the Electronic Structure of Solids*. (Oxford University Press, Oxford, 2012).
- [38] L. W. Wang and A. Zunger, “Solving Schroedinger’s equation around a desired energy: Application to silicon quantum dots,” *J. Chem. Phys.* **100**, 2394 (1994).
- [39] A. Canning, L. W. Wang, A. Williamson, and A. Zunger, “Parallel empirical pseudopotential electronic structure calculations for million atom systems,” *J. Comput. Phys.* **160**, 29-41 (2000).
- [40] R. Ketzmerick, G. Petschel, and T. Geisel, “Slow decay of temporal correlations in quantum systems with Cantor spectra,” *Phys. Rev. Lett.* **69**, 695 (1992).
- [41] M. Benzi, P. Boito, “Quadrature rule-based bounds for functions of adjacency matrices,” *Lin. Algebra Appl.* **433**, 637-652 (2010).
- [42] M. Benzi, and G. H. Golub, “Bounds for the entries of matrix functions with application to preconditioning,” *BIT* **39**, 417-438 (1999).
- [43] M. Benzi, and P. Boito, “Quadrature rule-based bounds for functions of adjacency matrices,” *Lin. Algebra Appl.* **433**, 637-652 (2010).
- [44] G. H. Golub, and G. Meurant, *Matrices, Moments and Quadrature with Applications*. (Princeton University Press, Princeton, NJ 2010).
- [45] L. Collatz, and U. Sinogowitz, “Spektren Endlicher Grafen,” *Abh. Math. Sem. Univ. Hamburg* **21**, 63-77 (1957).
- [46] P. Erdős, A. Rényi, “On Random Graphs,” I, *Publicationes Mathematicae* **6**, 290-297 (1959).
- [47] E. N. Gilbert, “Random Graphs,” *Annals Math. Stat.* **30**, 1141-1144 (1959).
- [48] A.-L. Barabási, R. Albert, “Emergence of scaling in random networks,” *Science* **286**, 509-512 (1999).
- [49] E. Wigner, “Characteristic vectors of bordered matrices with infinite dimensions,” *Ann. of Math.* **62**, 548-564 (1955).
- [50] I.J Farkas, I. Derényi, A.-L. Barabási, T. Vicsek, “Spectra of “real-world” graphs: Beyond the semicircle law.” *Phys. Rev. E* **64**, 026704 (2001).
- [51] M. Krivelevich, and B. Sudakov, “The largest eigenvalue of sparse random graphs,” *Combin. Probab. Comput.* **12**, 61-72 (2003).
- [52] R. Milo, S. Shen-Orr, S. Itzkovitz, et al., “Network motifs: simple building blocks of complex networks,” *Science* **298**, 824-827 (2002).
- [53] O. Sporns, and R. Kötter, “Motifs in brain networks,” *PLoS Biology* **2**, e369 (2004).
- [54] P. Yodzis, “Local trophodynamics and the interaction of marine mammals and fisheries in the Benguela ecosystem,” *J. Anim. Ecol.* **67**, 635-658 (1998).
- [55] G. A. Polis, “Complex trophic interactions in deserts: an empirical critique of food-web theory,” *Am. Nat.* **138**, 123-155 (1991).
- [56] C. Townsend, R. M. Thompson, and A. R. McIntosh, et al. “Disturbance, resource supply, and food-web architecture in streams,” *Ecol. Lett.* **1**, 200 (1998).
- [57] R. R. Christian, and J. J. Luczkovich, “Organizing and understanding a winter’s seagrass foodweb network through effective trophic levels,” *Ecol. Model.* **117**, 99-124 (1999).
- [58] P. H. Warren, “Spatial and temporal variation in the structure of a fresh-water food web,” *Oikos* **55**, 299-311 (1989).
- [59] R. B. Waide, and W. B. Reagan, (Eds.) *The Food Web of a Tropical Rainforest*. (University Chicago Press, Chicago, 1996).
- [60] N. D. Martinez, B. A. Hawkins, H. A. Dawah, and B. P. Feifarek, “Effects of sampling efforts on characterization of food web structure,” *Ecology* **80**, 1044-1055 (1999).
- [61] K. Havens, “Scale and structure in natural food webs,” *Science* **257**, 1107-1109 (1992).
- [62] S. Opitz, “Trophic Interactions in Caribbean coral reefs,” *ICLARM Tech. Rep.* 43, Manila, Philippines, (1996).
- [63] J. Memmott, N. D. Martinez, and J. E. Cohen, “Predators, parasites and pathogens: species richness, trophic generality, and body sizes in a natural food web,” *J. Animal Ecol.* **69**, 1-15 (2000).
- [64] J. Link, “Does food web theory work for marine ecosystems?,” *Mar. Ecol. Prog. Ser.* **230**, 1-9 (2002).
- [65] P. Yodzis, “Diffuse effects in food webs,” *Ecology* **81**, 261-266 (2000).
- [66] L. Goldwasser, and J. A. Roughgarden, “Construction and analysis of a large Caribbean food web,” *Ecology* **74**, 1216-1233 (1993).
- [67] N. D. Martinez, “Artifacts or attributes? Effects of resolution on the Little Rock Lake food web,” *Ecol. Monogr.* **61**, 367-392 (1991).
- [68] D. Baird, and R. E. Ulanowicz, “The seasonal dynamics of the Chesapeake Bay ecosystem,” *Ecol. Mon.* **59**, 329-364 (1989).
- [69] M. Huxman, S. Beany, and D. Raffaelli, “Do parasites reduce the chances of triangulation in a real food web?,” *Oikos* **76**, 284-300 (1996).

- [70] S. J. Hall, and D. Rafaelli, "Food-web patterns - lessons from a species-rich web," *J. Anim. Ecol.* **60**, 823-842 (1991).
- [71] A. Perna, S. Valverde, and J. Gautrais, et al. "Topological efficiency in the three-dimensional gallery networks of termite nests," *Physica A* **387**, 6235-6244 (2008).
- [72] J. Buhl, J. Gautrais, and R. V. Solé, et al. "Efficiency and robustness in ant networks of galleries," *Eur. Phys. J. B* **42**, 123-129 (2004).
- [73] D., Lusseau, "The emergent properties of a dolphin social network," *Proc. R. Soc. Lond. B (Suppl.)* **270**, 186-188 (2003).
- [74] N. P. Hummon, P. Doreian, and L. C. Freeman, "Analyzing the structure of the centrality-productivity literature created between 1948 and 1979," *Know.-Creat. Diffus. Util.* **11**, 459-480 (1990).
- [75] W. de Nooy, A. Mrvar, and V. Batagelj, *Exploratory Social Network Analysis with Pajek*, Cambridge University Press, Cambridge (2005).
- [76] V. Batagelj, and A. Mrvar, "Graph Drawing Contest 2001," <http://vlado.fmf.uni-lj.si/pub/GD/GD01.htm>. (2001).
- [77] "ODLIS: Online Dictionary of Library and Information Science," <http://vax.wcsu.edu/library/odlis.html>. (2002).
- [78] Roget's Thesaurus of English Words and Phrases, Project Gutenberg. <http://gutenberg.net/etext/22> (2002).
- [79] V. Batagelj, and A. Mrvar, Pajek datasets. Available at: <http://vlado.fmf.uni-lj.si/pub/networks/data/> (2006).
- [80] P. Uetz, Y.-A. Dong, and Ch. Zeretzke, et al., "Herpesviral protein networks and their interaction with the human proteome," *Science* **311**, 239-242 (2006).
- [81] D. LaCount, M. Vignali, and R. Chettier, et al. "A protein interaction network of the malaria parasite *Plasmodium falciparum*," *Nature* **438**, 103-107 (2005).
- [82] D. Bu, Y. Zhao, and L. Cai, et al. "Topological structure analysis of the protein-protein interaction network in budding yeast," *Nucleic Acids Res.* **31**, 2443-2450 (2003).
- [83] C. von Mering, R. Krause, and B. Snel, et al. "Comparative assessment of large-scale data sets of protein-protein interactions," *Nature* **417**, 399-403 (2002).
- [84] M. Motz, I. Kober, and C. Girardot, et al. "Elucidation of an Archaeal Replication Protein Network to Generate Enhanced PCR Enzymes," *J. Biol. Chem.* **277**, 16179-16188 (2002).
- [85] C. Y. Lin, C. L. Chen, C. S. Cho, et al. "hp-DPI: *Helicobacter pylori* database of protein interactomes. A combined experimental and inferring interactions," *Bioinformatics* **21**, 1288-1290 (2005).
- [86] G. Bultland, J. M. Peregrín-Alvarez, and J. Li, et al. "Interaction network containing conserved and essential protein complexes in *Escherichia coli*," *Nature* **433**, 531-537 (2005).
- [87] P. Noirot, and N. F. Noirot-Gross, "Protein interaction networks in bacteria," *Curr. Op. Microb.* **7**, 505-512 (2004).
- [88] R. Milo, S. Itzkovitz, and N. Kashtan, et al. "Superfamilies of evolved and designed networks," *Science* **303**, 1538-1542 (2004).
- [89] S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, "Network motifs in the transcriptional regulation network of *Escherichia coli*," *Nature Gen.* **31**, 64-68 (2002).
- [90] G. F. Davis, M. Yoo, and W. E. Baker, "The Small World of the American Corporate Elite, 1982-2001," *Strategic Organization* **1**, 301-326 (2003).
- [91] D. MacRae, "Direct factor analysis of sociometric data," *Sociometry* **23**, 360-371 (1960).
- [92] J. Moody, Data for this project was provided in part by NIH grants DA12831 and HD41877, those interested in obtaining a copy of these data should contact James Moody (moody.77@sociology.osu.edu), (2001).
- [93] W. Zachary, "An information flow model for conflict and fission in small groups," *J. Anthropol. Res.* **33**, 452-473 (1977).
- [94] L. D. Zeleny, "Adaptation of research findings in social leadership to college classroom procedures," *Sociometry* **13**, 314-328 (1950).
- [95] J. J. Potterat, L. Philips-Plummer, and S. Q. Muth, et al. "Risk network structure in the early epidemic phase of HIV transmission in Colorado Springs," *Sex. Transm. Infect.* **78**, i159-i163 (2002).
- [96] D. Krackhardt, "The ties that torture: Simmelian tie analysis in organizations" *Res. Sociol. Org.* **16**, 183-210 (1999).
- [97] J. H. Michael, and J. G. Massey, "Modeling the communication network in a sawmill" *Forest Prod. J.* **47**, 25-30 (1997).
- [98] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the internet topology" *Comp. Comm. Rev.* **29**, 251-262 (1999).
- [99] D. J. Watts, and S. H. Strogatz, "Collective dynamics of small-world networks" *Nature* **393**, 440-442 (1998).
- [100] C. R. Myers, "Software systems as complex networks: Structure, function, and evolvability of software collaboration graphs," *Phys. Rev. E* **68**, 046116 (2003)
- [101] M. E. J. Newman, "Assortative mixing in networks," *Phys. Rev. Lett.* **89**, 208701 (2002).