PHYSICAL REVIEW E **82**, 066102 (2010)

Quantifying network heterogeneity

Ernesto Estrada
Department of Mathematics & Statistics, Department of Physics, SUPA and Institute of Complex Systems at Strathclyde, University of Strathclyde, Glasgow G1 1XH, United Kingdom
(Received 4 October 2010; published 2 December 2010)

Despite degree distributions give some insights about how heterogeneous a network is, they fail in giving a unique quantitative characterization of network heterogeneity. This is particularly the case when several different distributions fit for the same network, when the number of data points is very scarce due to network size, or when we have to compare two networks with completely different degree distributions. Here we propose a unique characterization of network heterogeneity based on the difference of functions of node degrees for all pairs of linked nodes. We show that this heterogeneity index can be expressed as a quadratic form of the Laplacian matrix of the network, which allows a spectral representation of network heterogeneity. We give bounds for this index, which is equal to zero for any regular network and equal to one only for star graphs. Using it we study random networks showing that those generated by the Erdös-Rényi algorithm have zero heterogeneity, and those generated by the preferential attachment method of Barabási and Albert display only 11% of the heterogeneity of a star graph. We finally study 52 real-world networks and we found that they display a large variety of heterogeneities. We also show that a classification system based on degree distributions does not reflect the heterogeneity properties of real-world networks.

DOI: 10.1103/PhysRevE.82.066102 PACS number(s): 89.75.Fb, 89.75.Hc, 02.10.Ox

I. INTRODUCTION

Complex networks are the structural skeleton of biological, ecological, technological, and socioeconomic systems [1,2]. They are formed by a set of nodes V representing the entities of these systems and a set of links E representing relationships between pairs of nodes [3,4]. Despite the disparate nature of the systems represented by these networks they share several universal topological properties, such as *small worldness* [5], *scale freeness* [6], the existence of network motifs [7], and self-similarity characteristics [8]. A great deal of attention has been paid to the scale-free property shared by many real-world networks, which contrasts with the regularity observed in random network models like the one proposed by Erdös and Rényi (ER) [9]. A network is said to have scale-free properties if it displays a power-law degree distribution. That is, let $p(k)=n(k)/n$ be the probability of randomly selecting a node of degree k in a network, where $n(k)$ is the number of nodes having degree k in a network of size $|V|=n$. The degree is the number of links incident with the corresponding node. Then, a plot of $p(k)$ versus k represents the degree distribution for the network [3,4]. A random ER network displays a Poisson degree distribution. However, many real-world networks have been observed to display degree distributions in which the probability $p(k)$ decays as a power-law with the degree $k: p(k) \sim k^{-\gamma}$.

In a regular network all nodes have exactly the same degree k. Then, we can consider that networks with Poissonian degree distributions are almost regular in the sense that most of their nodes have degree about $\bar{k}$, and few nodes have very small or very high degree. However, in a scale-free network the deviations from this regularity are very high. For instance, the probability of finding nodes with very small degree is very high, e.g., $p(1)\sim 1$, but very few nodes have very large degree, e.g., for $\gamma > 1$, $p(100) < 0.01$. This situation resembles the one observed for the star graph in which one node has degree $n-1$ and the rest of the nodes have degree one. The star graph has average degree $\bar{k}(S_n)=\frac{2(n-1)}{n}$, which tends to 2 for very large number of nodes, i.e., $\bar{k}(S_n) \to 2$ as $n\to\infty$. In both, star graphs and scale-free networks, there are nodes with degree which are significantly larger than the average degree of the network. This structural characteristic reflects their "irregularity" or as it is known in the network literature their "heterogeneity" [3,4].

The problem of quantifying network heterogeneity arises from the following situation. We know qualitatively that a scale-free network is more heterogeneous than a Poissonian network, and we also know how to compare the heterogeneity of two networks with power-law degree distributions. However, there are many networks that do not display power-law degree distributions, but any of the many fat-tail distributions such as lognormal, Burr, Gamma, stretched exponential, etc. For instance, Stumpf and Ingram [10] analyzed the empirical data for several protein-protein interaction networks (PPIs). The PPIs of yeast, *H. pylori* and *E. coli* were better fitted to a stretched exponential (Weibull distribution), the PPI of *C. elegans* displayed a power-law degree distribution and that of *D. melanogaster* was fitted to a gamma distribution. How can we rank them according to their heterogeneity? The obvious answer is that we need an index that uniquely accounts for the heterogeneity of any network irrespective of its degree distribution.

In addition, the identification of which degree distribution a network has, is not a trivial question. There are hundreds of possible distributions to tests. Sometimes the differences in the statistical fittings between several distributions are very small. More difficult is the situation for relatively small networks where the number of data points is not enough for having a good fit of any of the candidate distributions. Consequently, having a heterogeneity index, which is minimized for regular networks and maximized for starlike ones, is an urgent necessity.

II. EARLY NETWORK IRREGULARITY INDICES

In order to place our current research in context we describe here briefly some early attempts in the mathematical literature to characterize the irregularity of a graph. The main message of this section is that those early indices fail in identifying starlike graphs as the most heterogeneous ones. We start by describing an index based on the fact that the irregularity of a network can be intuitively accounted for by the difference of the node degree and the average degree. This index was proposed by Bell in 1992 as the variance of node degrees [11],

$$\mathrm{VAR} = \frac{1}{n}\sum_{i=1}^{n}(k_i - \bar{k})^2. \tag{1}$$

It has to be said, however, that this index was first introduced as a measure of centralization in social networks by Snijders [12]. Bell's index is minimized for regular graphs as expected from the fact that there is no heterogeneity at all in their degrees. However, the maximum value is not obtained for star graphs but depends on the number of nodes in the network (see for instance p. 181 on Ref. [13]). That is, for different number of nodes, the structures of two networks displaying the maximum variances of their respective node degrees are not necessarily similar. For details on the structure of the networks that maximize this index the reader is referred to Ref. [12].

In a historical context, the study of network irregularity was first proposed in the seminal paper of Collatz and Sinogowitz in 1957 [14], where they proposed the following index of irregularity:

$$\mathrm{CS}(G) = \lambda_1 - \bar{k}, \tag{2}$$

where λ_1 is the principal eigenvalue of the adjacency matrix and $\bar{k}$ is the average degree. This index is also zero for regular graphs and Collatz and Sinogowitz [14] conjectured that it is maximized by star graphs. Despite this is the case for graphs with up to five nodes, it has been found that the CS index is maximized for several families of graphs different from star graphs [15]. Other approaches to quantify graph irregularity were conceptualized by Chartrand, Erdös, and Oellermann [16]. Albertson later recognized that these attempts do not capture the irregularity of a graph in a single parameter and proposed yet another index which is defined as [17]

$$A(G) = \sum_{(i,j)\in E}|k_i - k_j|. \tag{3}$$

This idea is very close to the one we are going to develop in the current work, which is that of basing the irregularity index on a local measure of irregularity, e.g., the value $|k_i - k_j|$ for the link $e_{i,j}$. However, here again this index fails in quantifying correctly what we intuitively consider as a heterogeneous network. Despite this index is minimized for regular graphs, it is not always maximized for star graphs. In fact, this index is maximized for a particular class of complicated graphs consisting of a clique, an independent set, and some links joining a node in the clique to another one in the independent set [18]. In closing, these irregularity indices do not reflect our intuition that the starlike graphs are the most heterogeneous ones. Consequently, we propose in the next section an index which accounts for the heterogeneity of a network having the minimum value for any regular graph and being maximized only for stars.

III. NETWORK HETEROGENEITY INDEX

We start by defining a local index that accounts for the irregularity of a single link. Let $i,j\in E$, then we define the irregularity of the i,j-link as

$$I_{ij} = [f(k_i) - f(k_j)]^2, \tag{4}$$

where $f(k_i)$ is a function of the node degree. For the sake of mathematical convenience as we will see later we select here $f(k_i)=k_i^{-1/2}$. This function takes the value of zero if the two nodes have the same degree as it happens in regular networks and it is maximized when the difference of both degrees increases. For instance, let us consider a node of degree one connected to a node of degree k, then as $k\to\infty$, $I_{pq}\to 1$. In addition, this function also accounts for the relative difference between the degrees of the two nodes. For instance, if we have two links $e_{p,q}$ and $e_{r,s}$, such that $k_p=10$, $k_q=2$, $k_r=100$, and $k_s=92$, the simple difference of node degrees does not distinguish between the two links, i.e., $[k_p-k_q]^2=[k_r-k_s]^2=64$. However, using $f(k_i)=k_i^{-1/2}$ we have, $I_{pq}=0.153$ and $I_{rs}=1.81\cdot 10^{-5}$, which indicates the relatively larger irregularity of the first link compared to the second.

Then, the heterogeneity index that we propose here is simply defined as the sum of the link irregularity for all links in the network,

$$\rho'(G) = \sum_{i,j\in E}(k_i^{-1/2} - k_j^{-1/2})^2. \tag{5}$$

For regular networks this quantity is equal to zero. However, as the difference in the degrees of adjacent nodes increases the index also increases. The main advantage of defining this index as the sum of square differences of a function of node degrees is that we can express it in terms of a quadratic form of the Laplacian matrix of the network (see Fact 8.15.1 on p. 550 of Ref. [19]). That is, let us define the Laplacian matrix $\mathbf{L}$ of a network as $\mathbf{L}=\mathbf{K}-\mathbf{A}$, where $\mathbf{K}$ is a diagonal matrix of degrees and $\mathbf{A}$ is the adjacency matrix of the network, whose A_{ij} is one if, and only if, the corresponding nodes are joined by a link or zero otherwise. The entries of $\mathbf{L}$ are then given by

$$L_{ij} = \begin{cases} k_i & \text{for } i=j, \\ -1 & \text{for } i\sim j, \\ 0 & \text{otherwise}, \end{cases}$$

where $i\sim j$ stands for pairs of adjacent nodes.

The function $z(x)=\sum_{i,j\in E}(x_i-x_j)^2 w_{ij}$, where x_i are values assigned to the nodes, w_{ij} are weights for the links, and the vector of x_i entries has the constraint expressed by $\sum_{i,j\in E}x_i x_j m_{ij}=1$, where m_{ij} are the elements of a given symmetric matrix $\mathbf{M}$, was previously used by Capocci *et al.* [20] for detecting communities in large networks. Then, the stationary points of z over all x were found as solutions of $(\mathbf{D}-\mathbf{W})\mathbf{x}=\mu\mathbf{M}\mathbf{x}$, where $(\mathbf{D}-\mathbf{W})$ is the weighted Laplacian matrix and μ are Lagrange multipliers. If we consider nonweighted networks and take x_i as the inverse square root of the node degree then we have that the two apparently disconnected methods are equivalent, despite one is used for detecting communities in complex networks and the other for quantifying network heterogeneity.

Let $|\mathbf{k}^{-1/2}\rangle=(k_1^{-1/2},k_2^{-1/2},\ldots,k_n^{-1/2})$ represent a column vector where k_i is the degree of the node i. Then, it is easy to realize that expression (5) can be stated as a quadratic form of the Laplacian matrix,

$$\rho'(G)=\sum_{i,j\in E}(k_i^{-1/2}-k_j^{-1/2})^2=\frac{1}{2}\langle\mathbf{k}^{-1/2}|\mathbf{L}|\mathbf{k}^{-1/2}\rangle$$
$$=n-2\sum_{i,j\in E}(k_ik_j)^{-1/2}. \tag{6}$$

The second term in the right hand side of Eq. (6) is two times the so-called *Randić index* ${}^1R_{-1/2}$ [21] of the network, which has been extensively studied in the mathematical literature [22,23]. Then

$$\rho'(G)=n-2\,{}^1R_{-1/2}. \tag{7}$$

The advantage of using $f(k_i)=k_i^{-1/2}$ is that expression (7) depends only on the Randić index, for which the extremal graphs are well known to coincide with the ones determining the two extremes of irregularity in which we are interested here. That is, for connected networks the Randić index is bounded as follows [23]:

$$\sqrt{n-1}\le{}^1R_{-1/2}\le\frac{n}{2}, \tag{8}$$

where the lower bound is attained for the star S_n and the upper bound is attained for any regular network with n nodes. Then, we can define the normalized heterogeneity index $\rho(G)$ as

$$\rho(G)=\frac{n-2\,{}^1R_{-1/2}}{n-2\sqrt{n-1}}, \tag{9}$$

which is zero for any regular network and one for the star graph, i.e., $0\le\rho(G)\le 1$. Then, heterogeneous starlike networks are expected to have values of $\rho(G)$ close to one. On the other hand, more regular networks are expected to have values close to zero. The normalized heterogeneity index can be simply written as follows:

$$\rho(G)=\frac{\sum_{i,j\in E}(k_i^{-1/2}-k_j^{-1/2})^2}{n-2\sqrt{n-1}}. \tag{10}$$

IV. SPECTRAL REPRESENTATION

Let $\vec{\varphi}_j$ be an orthonormal eigenvector of the Laplacian matrix associated with the μ_j eigenvalue. We recall that the Laplacian matrix is positive semidefinite, which for a connected network means that: $0=\mu_1<\mu_2\le\ldots\le\mu_n$. Let

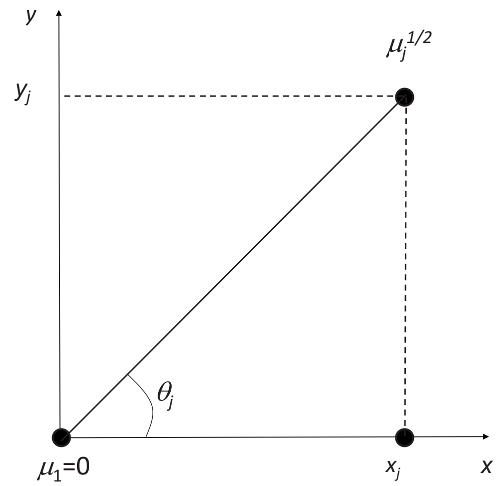

FIG. 1. Illustration of the projection of $\sqrt{\mu_{j>1}}$ in terms of its magnitude and the angle θ_j formed between an orthonormal eigenvector associated to $\mu_{j>1}$ and the vector $\mathbf{k}^{-1/2}$.

$$\cos\theta_j=\frac{\mathbf{k}^{-1/2}\cdot\vec{\varphi}_j}{\|\mathbf{k}^{-1/2}\|}, \tag{11}$$

be the angle between the orthonormal eigenvector $\vec{\varphi}_j$ and the vector $\mathbf{k}^{-1/2}$ previously defined. The Euclidean norm $\|\mathbf{k}^{-1/2}\|$ can be written as $\|\mathbf{k}^{-1/2}\|=\sqrt{\Sigma_i k_i^{-1}}=\sqrt{{}^0R_{-1}}$. Then, using the Euler theorem (see p. 457 on Ref. [24]) the *Randić index* can be expressed as follows:

$${}^1R_{-1/2}=\frac{1}{2}\left[n-\frac{1}{{}^0R_{-1}}\sum_{j=2}^{n}\mu_j\cos^2\theta_j\right]. \tag{12}$$

The term $\cos^2\theta_j$ represents the "contribution" of the normalized degree to the corresponding eigenvector (or vice versa). For instance, $\cos^2\theta_j=0$ means that the vector $\mathbf{k}^{-1/2}$ is perpendicular to the corresponding eigenvector, and no "duplicated" information is contained in both vectors.

Now let us consider that the eigenvalue $\mu_1=0$ represents the origin of a Cartesian coordinate system. Let us represent $\sqrt{\mu_{j>1}}$ as a point in this system with coordinates given by the magnitude of $\sqrt{\mu_{j>1}}$ and the angle θ_j formed between an orthonormal eigenvector associated to $\mu_{j>1}$ and the vector $\mathbf{k}^{-1/2}$. Then, the projection of $\sqrt{\mu_{j>1}}$ on the x axis is given by $x_j=\sqrt{\mu_{j>1}}\cos\theta_j$, and the projection of $\sqrt{\mu_{j>1}}$ on the y axis is given by $y_j=\sqrt{\mu_{j>1}}\sin\theta_j$.

This means that the heterogeneity index $\rho(G)$ can be written as

$$\rho(G)=\frac{{}^0R_{-1}}{n-2\sqrt{n-1}}\sum_{j=1}^{n}x_j^2. \tag{13}$$

This scenario is equivalent to considering that x_j is the adjacent cathetus of the triangle formed by the points $\mu_1=0$, $\sqrt{\mu_{j>1}}$ and x_j (see Fig. 1). Then, $\rho(G)$ can be interpreted as the sum of the squares of adjacent catheti for all triangles formed by the spectral projection of the Laplacian eigenvalues. Consequently, we can represent a network in a graphical form by plotting x_j vs y_j for all values of j, where the heterogeneity is given by the sum of the squares of the projections of all these points on the abscissa. Obviously, all projections on y axis are positive but those on x axis can have positive and negative signs. We will call these plots heterogeneity plots or simply H plots.

V. HETEROGENEITY IN RANDOM NETWORKS

We start our analysis by considering the Erdös-Rényi random networks $G_{n,p}$ generated by taking n nodes which are linked by pairs according to a probability p, $0 \leq p \leq 1$. $G_{n,p}$ is almost surely connected when $p \sim \omega(n)\log n/n$, where $\omega(n) \to \infty$ [25]. In this limit, the degrees of almost all nodes are asymptotically equal [25], which means that $\rho(G) \approx 0$.

In the case of scale-free networks we have studied empirically the networks generated with the preferential attachment algorithm of Barabási and Albert (BA) [6]. By studying networks having 5000; 10 000, and 20 000 nodes and average degrees ranging from 2 to 16 we found that the heterogeneity of BA networks decays as a power law of the form $\rho(BA) \approx a\bar{k}^{-\alpha}+b$, where the parameters of the best fits obtained are given in Table I together with the Pearson correlation coefficient r.

Then, using the whole data set of points we have found that the heterogeneity of BA networks can be estimated in general by using the following expression obtained by using nonlinear fit and displaying correlation coefficient of $r = 0.9983$,

$$\rho(BA) \approx 0.2889\bar{k}^{-0.9012} + 0.1109. \tag{14}$$

This means that as the average degree increases the heterogeneity of BA networks tends to a constant value, $\rho(BA) \approx 0.1109$. Consequently, the BA model is a poor generator of network heterogeneity as it is able to produce only about 11% of the heterogeneity of a starlike graph.

Another interesting characteristic of the heterogeneity of random networks is given by their spectral representations. It has been previously observed that the Laplacian eigenvalue distributions share some similarities with the degree distributions of ER and BA networks. That is, the Laplacian eigenvalue of an ER has a Poisson-like distribution, while that of a BA network has a power-law tail. In general, Zhan *et al.* [26] have observed that for ER and BA networks the Laplacian eigenvalue curves are very similar to their node degree curves. This characteristic has an important impact in the spectral representation of heterogeneity for ER and BA networks.

In Fig. 2 we illustrate the spectral heterogeneity H plots for two ER and two BA networks with average degrees 8 (left) and 16 (right). The values on the x axis are normalized between -1 and 1, and those of the y axis between 0 and 2 to have similar length in both scales. Both types of networks display a characteristic plot, which we have observed for all networks generated with these models. In the case of ER networks where the Laplacian eigenvalues are distributed according to a Poissonian law the plot of x_j vs y_j is characterized by a regular distribution of the points with an almost squared shape. In the case of BA networks the plots have a characteristic V shape, which is very narrow for values around the point (0,0). We have seen this shape in many other networks with fat-tail degree distributions as we will see in the next section. The exact analytical expressions for these plots are not explored in the current work and we think that its exploration deserves further attention.

TABLE I. Parameters for the power-law relations between heterogeneity and average degree in ER networks.

n	a	b	α	r
5000	0.2766	0.1031	0.8017	0.9982
10000	0.3001	0.1184	1.0181	0.9986
20000	0.2939	0.1096	0.8898	0.9998

VI. HETEROGENEITY IN REAL-WORLD NETWORKS

We study here three groups of networks loosely classified according to their degree distributions into the classes of (i) homogeneous, (ii) exponential, and (iii) fat-tailed networks. The first group is formed by networks characterized by having Poisson, Gaussian, or Uniform degree distributions. The second group is formed by networks having exponential-like distributions and the third group is formed by those having power-law or other heavy-tailed distributions. There are 16, 18, and 18 networks in each group and they cover social, ecological, technological, biological, and informational systems. Networks in group (i) include the food webs of Benguela, Coachella Valley, Reef Small, Shelf, Skipwith pond, St. Marks seagrass, and Stony stream; the social networks of corporate elite in USA, inmates in prison, the friendship network between physicians (Galesburg), the friendship ties among the employees in a small hi-tech computer firm which sells, installs, and maintains computer systems (high-tech), and a sawmill communication network; three electronic sequential logic circuits parsed from the ISCAS89 benchmark set, and a network of the Roget thesaurus. Networks in group (ii) include the food webs of Bridge Brook, Canton Creek, Chesapeake Bay, El Verde rainforest, Little Rock, St. Martin, and Ythan estuary with and without parasites; the social networks of injecting drug users, a social network among college students in a course about leadership and the Zachary karate club; the neural network of *C. elegans*; the western USA power grid; a citation network consisting of papers published in the *Proceedings of Graph Drawing* in the period 1994–2000 (GD); the software network of XMMS; the USA airport transportation network of 1997 and the PPI network of *D. melanogaster*. The third group is formed by: the social networks of persons with HIV infection during its early epidemic phase in Colorado Springs, a scientific collaboration network in the field of computational geometry, and two sexual networks, one consisting of heterosexual relations only and the other including both heterosexual and homosexual relationships; two versions of Internet at autonomous system of 1997 and 1998; a semantic network of the Online Dictionary of Library and Information Science (ODLIS); the food webs of Scotch Broom and Grassland; a citation network in the field of "small-world;" the software networks of

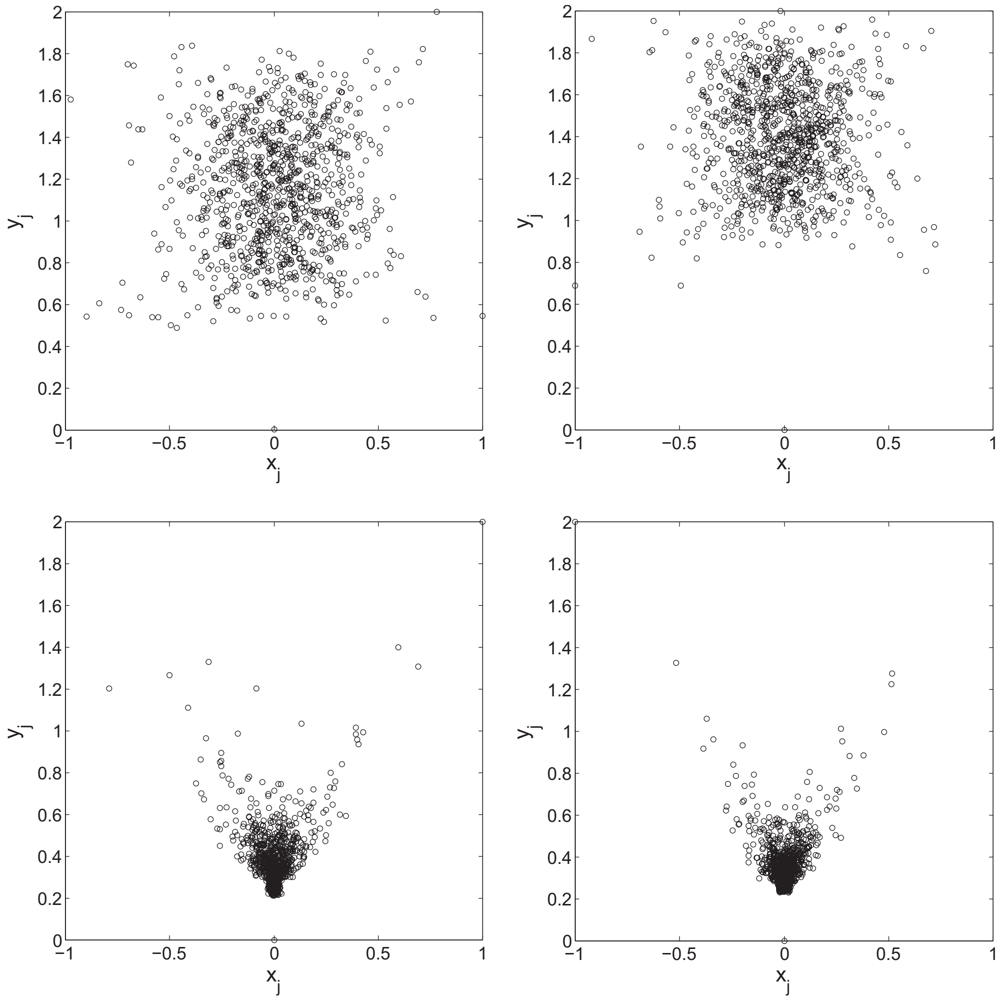

FIG. 2. Illustration of the H plots for random networks obtained by the Erdös-Rényi (top) and Barabási-Albert (bottom) methods. Networks on the left hand side have average degree equal to 8 and those on the right have average degree equal to 16. All networks have 1000 nodes.

Abi, Digital, MySQL, and VTK; the transcription networks of yeast, *E. coli* and urchins; the PPI networks of human and yeast. Information about these networks as well as the appropriate references is given in [27,28].

The first general characteristic that we observe for this wide group of networks is that they are not very heterogeneous in the sense of being starlike. The average heterogeneity for these networks is $\bar{\rho}(G)=0.218\pm0.129$. In fact the most heterogeneous network among the ones studied here is the 1997 version of Internet at the autonomous system, which has $\rho(G)=0.548$. This means that this network has 55% of the heterogeneity of a star graph of the same size. The analysis of the three groups reveals the following statistics. For group (i) $\bar{\rho}(G)=0.088\pm0.052$, group (ii) $\bar{\rho}(G)=0.218\pm0.081$, group (iii) $\bar{\rho}(G)=0.340\pm0.097$. The largest relative deviations are observed for the group we have called "homogeneous" for which the values range from 0.04 to 0.24. In Fig. 3 we illustrate the values of heterogeneity for all networks in the three groups. It can be seen that the three groups are overlapped to each other indicating that a classification of networks into broad groups according to their degree distributions does not reflect the real heterogeneity of the networks.

A different way of grouping these networks is according to their values of the heterogeneity index indistinctly of their degree distributions. If we allow for two groups only, let say homogeneous and "heterogeneous" networks we obtain two clusters of networks which are centered at $\bar{\rho}_1=0.092\pm0.038$ and $\bar{\rho}_2=0.310\pm0.088$, respectively. This clustering is obtained by using the K-mean clustering technique. The first group is now formed by 22 networks, 73% of which display Poisson, Gaussian or Uniform degree distributions and the rest displays exponential-like ones. Then, the second group is formed by 30 networks, 60% of which display some kind of fat-tail degree distribution, 37% display exponential-like and only one network displays a uniform-like distribution. This second group can be further subdivided into two subgroups accounting for moderately heterogeneous and highly heterogeneous networks. Both subgroups are centered at $\bar{\rho}_2(1)=0.269\pm0.036$ and $\bar{\rho}_2(2)=0.444\pm0.072$, respectively. The last subgroup includes only seven of the 52 networks studied here, all of them having fat-tail degree distribution but one that display an exponential-like one. These highly heterogeneous networks correspond to the USA airport network of 1997 (the one displaying exponential degree distribution), Internet versions

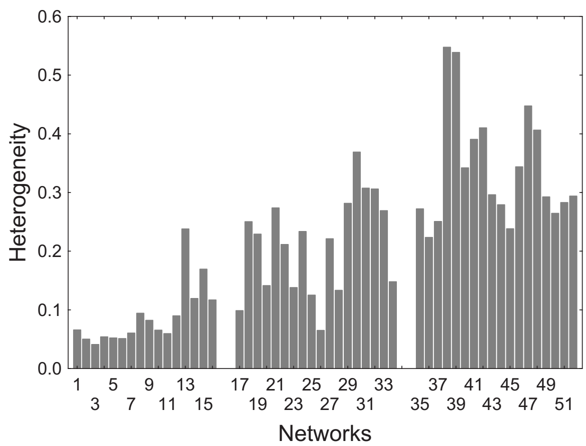

FIG. 3. Plot of the values of the heterogeneity index for 52 real-world networks grouped according to their general kind of degree distribution. The first group corresponds to networks with homogeneous (Poisson, Gaussian, or uniform), the second to those having exponential-like, and the third to those having fat-tail degree distribution.

of 1997 and 1998 at autonomous systems, a food web of Scotch Broom, the network of citations in the field of "small-world," and the transcription networks of yeast and *E. coli.* In closing, a classification system based only on the degree distribution does not reflect all characteristics of network heterogeneity.

We have seen in the previous section that H plots reflect in some way the characteristics of the degree distributions. For instance, we have seen that networks with Poisson degree distribution display a very regular distribution of the points in the H plot, while scale-free networks display a characteristic V-shape distribution of the points. In Fig. 4 we illustrate the H plots of six real-world networks, with homogeneous, exponential, and fat-tail degree distributions. It can be easily observed that H plots display some qualitative differences for networks with different kinds of degree distribution. For instance, networks with homogeneous degree distributions display H plots in which all points are regularly distributed inside a square. However, networks with exponential degree distributions display an accumulation of the points around the vertical line $x_j=0$, which means that most of the points are contained in a rectangle elongated along the y axis. Finally, networks with fat-tail degree distributions display the characteristic V-shape distributions of the points in the H plot.

This classification of networks is very loose and an exploration of the whole universe of complex network will display many intermediate cases which are difficult to distinguish qualitatively or even quantitatively. However, as an attempt to quantify the "shape" of the H plots as another characteristic feature of network heterogeneity we propose to use the ratio of the lengths of both sides of the rectangle containing most of the points in these plots. That is, let us consider a rectangle with center at $\bar{x}$, $\bar{y}$, where $\bar{x}$ and $\bar{y}$ are the averages of x_j and y_j, respectively. The lengths of the two sides of this rectangle containing most of the points in the H plot are given by the standard deviations of these points respect to $\bar{x}$ and $\bar{y}$, respectively,

$$l_x = 2\sqrt{\frac{1}{n}\sum_{j=1}^{n}(x_j-\bar{x})^2}, \tag{15}$$

$$l_y = 2\sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j-\bar{y})^2}. \tag{16}$$

Then, let $l_{\min}=\min(l_x,l_y)$ and $l_{\max}=\max(l_x,l_y)$. We define the ratio between the two lengths as $\Omega=l_{\min}/l_{\max}$, which is $\Omega=1$ if the points are distributed regularly on a square and $\Omega\to 0$ as the points concentrate around a line in the center of the H plot.

The values of Ω for random networks with Poisson degree distributions are close to one. This is also the case for the two networks displaying homogeneous degree distribution in Fig. 4. For instance, $\Omega=0.935$ for the network of inmates in prison, and $\Omega=0.757$ for the network of Roget thesaurus. However, for networks with exponential degree distributions these values are closer to zero. For instance, for the Bridge Brook food web $\Omega=0.480$ and for the network of drug injecting users it is: $\Omega=0.383$. Networks with fat-tail degree distributions are better recognized by the V shape in their H plots but they also tend to have values of the ratio parameter around 0.5, e.g., $\Omega=0.507$ for the transcription networks of yeast and $\Omega=0.490$ for Internet. However, these values cannot be taken as a rule of thumb in order to match H plots and degree distributions. In a similar way as for the heterogeneity index here also there is large overlap between the different classes. For instance, $\bar{\Omega}=0.72\pm0.16$, $\bar{\Omega}=0.61\pm0.13$, and $\bar{\Omega}=0.51\pm0.05$ for networks with homogeneous, exponential and fat-tail degree distributions, respectively, showing a large overlapping specially between the first two groups.

VII. CONCLUSIONS

We have defined here an index that accounts for the heterogeneity of a network by using the sum of differences of some function of the node degrees for linked pairs of nodes. This index is then expressed as a quadratic form of the Laplacian matrix of the network, also allowing a spectral representation on the basis of Laplacian eigenvalues and eigenvectors. For computational purposes it is easily computed from any of the following equalities:

$$\rho(G) = \frac{\sum_{i,j\in E}(k_i^{-1/2}-k_j^{-1/2})^2}{n-2\sqrt{n-1}} = \frac{\frac{1}{2}\langle \mathbf{k}^{-1/2}|\mathbf{L}|\mathbf{k}^{-1/2}\rangle}{n-2\sqrt{n-1}} = \frac{n-2\sum_{i,j\in E}(k_ik_j)^{-1/2}}{n-2\sqrt{n-1}}. \tag{17}$$

More general strategies for assessing the relevance of node features in networks were proposed by Bianconi *et al.* [29]. In that work, more general characteristics of nodes, such as age, gender, nationality, abundance of proteins in the cell, or geographical position of airports are considered. Here

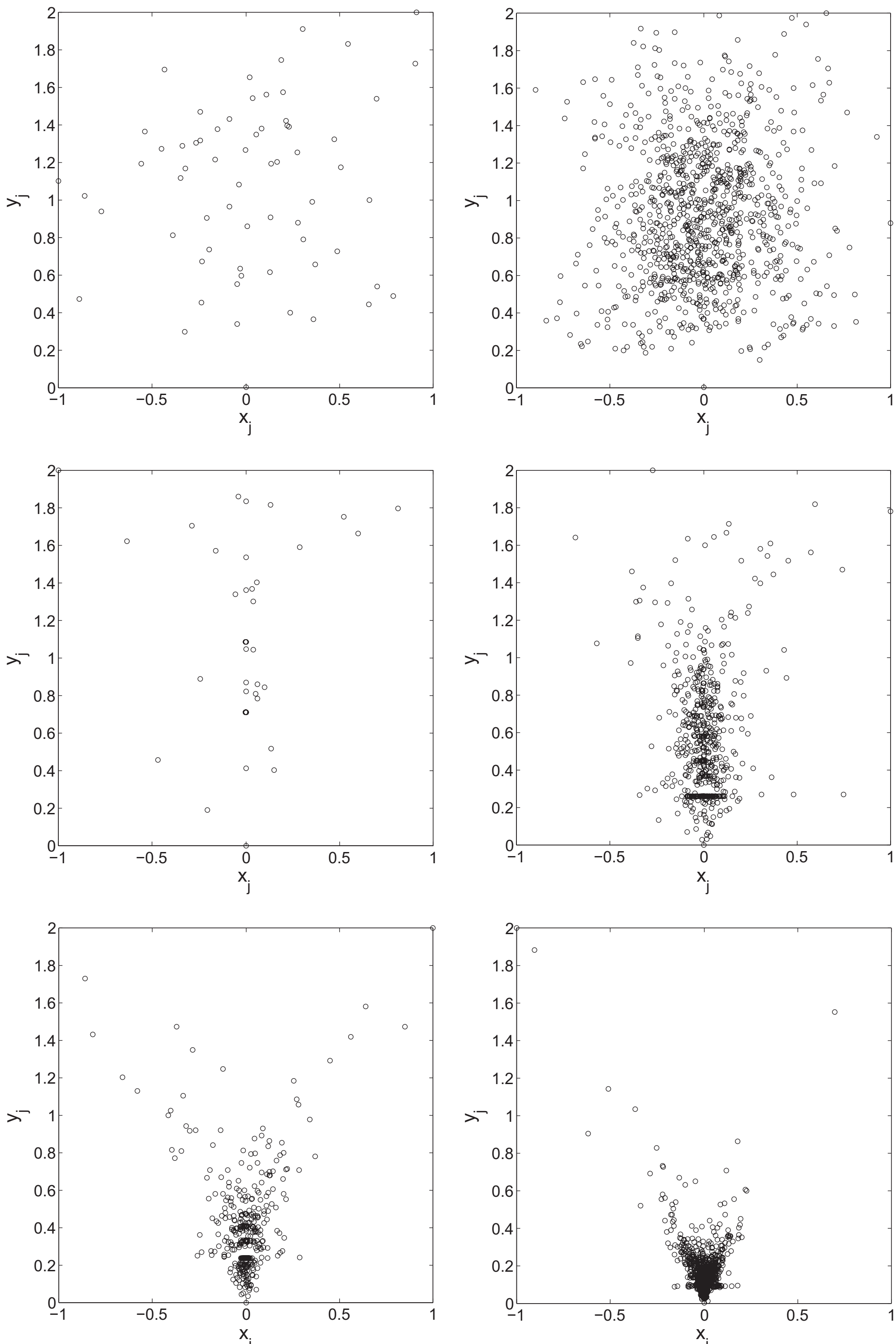

FIG. 4. H plots for six real-world networks from different scenarios. Those in the first line have homogeneous (Poisson, Gaussian, or uniform), those in the second line have exponential-like, and the ones in the third line display fat-tail degree distribution. Networks on the left hand side have relatively small sizes in comparison with those at the right hand side of the figure as seen by the number of points in the respective plots.

we concentrate on the heterogeneity that node degrees of linked nodes introduce to a given network. However, it is worth mentioning here that the possibility of extending this theoretical scheme to other "centrality" measures or nontopological characteristics of nodes like the ones previously mentioned is straightforward.

Using the spectral formula for the index introduced here, we have designed a way of representing the heterogeneity of a network by using a so-called H plot. These plots are build by representing the square root of every Laplacian eigenvalue $\sqrt{\mu_{j>1}}$ as a point in a system with coordinates given by the magnitude of $\sqrt{\mu_{j>1}}$ and the angle θ_j formed between an orthonormal eigenvector associated to $\mu_{j>1}$ and the vector $\mathbf{k}^{-1/2}$. We have shown that H plots for networks with Poisson degree distributions are characterized by a regular distribution of the points on an square with center at $\bar{x}$, $\bar{y}$ and side lengths determined by the standard deviation of the points from the mean. In case of networks with power-law degree distributions this plot displays a characteristic V shape, which distinguishes them clearly from the rest.

When we apply these heterogeneity tools to random and real-world networks we extract the following general conclusions. (i) As expected networks with Poisson degree distributions like the ones generated by Erdös-Rényi approach are very homogeneous, i.e., $\rho(G) \to 0$ for $n \to \infty$. (ii) Surprisingly, networks with power-law degree distributions generated by the Barabási-Albert preferential attachment method display very poor heterogeneity, which tends to a constant value $\rho(G) \to 0.12$ for large sizes and large average node degrees. (iii) Real-world networks display a large variety of heterogeneities ranging from values close to zero to values of about 0.55. No one network from a pool of 52 studied here displayed heterogeneity close to that of a star graph. The largest value being found so far corresponds to the Internet, which displays 55% of the heterogeneity of a star graph. (iv) A classification system based on degree distributions does not reflect the heterogeneity properties of real-world networks. (v) Attempts to use the heterogeneity indices defined here to infer the degree distribution of networks is in general not valid. Despite there is some match between heterogeneity indices and degree distributions there are several pathological cases which deviate from the general trends. However, networks with fat-tail degree distributions appear to be very well characterized by their heterogeneity index as well as by the characteristic shape of their H plots. We hope the current work helps researchers in different areas using network-based strategies to gain insights about the degree of heterogeneity that their networks have as well as relating it with other organizational and functional properties of networks.

ACKNOWLEDGMENT

Partial financial support from the Principal, University of Strathclyde through the New Professors' Fund is thanked.

[1] S. H. Strogatz, Nature (London) **410**, 268 (2001).
[2] *Network Science. Complexity in Nature and Technology*, edited by E. Estrada, M. Fox, D. Higham, and G.-L. Oppo (Springer, New York, 2010).
[3] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[4] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[5] D. J. Watts and H. Strogatz, Nature (London) **393**, 440 (1998).
[6] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[7] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science **298**, 824 (2002).
[8] C. Song, S. Havlin, and H. A. Makse, Nature (London) **433**, 392 (2005).
[9] P. Erdös and A. Rényi, Publ. Math. **6**, 290 (1959).
[10] M. P. H. Stumpf and P. J. Ingram, EPL **71**, 152 (2005).
[11] F. K. Bell, Linear Algebra Appl. **161**, 45 (1992).
[12] T. A. B. Snijders, Soc. Networks **3**, 163 (1981).
[13] S. Wasserman and K. Faust, *Social Network Analysis* (Cambridge University Press, Cambridge, England, 1994).
[14] L. Collatz and U. Sinogowitz, Abh. Math. Semin. Univ. Hambg. **21**, 63 (1957).
[15] D. Cvetković and P. Rowlinson, Publ. Inst. Math. (Beograd) **44**, 29 (1988).
[16] G. Chartrand, P. Erdös, and O. R. Oellermann, Coll. Math. J. **19**, 36 (1988).
[17] M. O. Albertson, Ars Comb. **46**, 219 (1997).
[18] P. Hansen and H. Mélot, *Graphs and Discovery*, edited by S. Fajtlowicz, DIMACS Series in Discrete Mathematics and Theoretical Computer Science (American Mathematical Society, Providence, 2005), Vol. 69, p. 253.
[19] D. S. Bernstein, *Matrix Mathematics. Theory, Facts and Formulas* (Princeton Academic Press, Princeton, 2009).
[20] A. Capocci, V. D. P. Servedio, G. Caldarelli, and F. Colaiori, Physica A **352**, 669 (2005).
[21] M. Randić, J. Am. Chem. Soc. **97**, 6609 (1975).
[22] B. Bollobàs and P. Erdös, Ars Comb. **50**, 225 (1998).
[23] X. Li and Y. Shi, MATCH Comm. Math. Compu. Chem. **59**, 127 (2008).
[24] F. E. Hohn, *Elementary Matrix Algebra* (Dover, New York, 1973).
[25] R. Durret, *Random Graph Dynamics* (Cambridge University Press, Cambridge, England, 2007).
[26] C. Zhan, G. Chen, and L. F. Yeung, Physica A **389**, 1779 (2010).
[27] E. Estrada, Eur. Phys. J. B **52**, 563 (2006).
[28] E. Estrada, D. J. Higham, and N. Hatano, Physica A **388**, 764 (2009).
[29] G. Bianconi, P. Pin, and M. Marsili, Proc. Natl. Acad. Sci. U.S.A. **106**, 11433 (2009).