PROJECT PERIODIC REPORT

Grant Agreement number: 288956

Project acronym: NADINE

Project title: New tools and Algorithms for Directed Network analysis

Funding Scheme: Small or medium-scale focused research project (STREP)

Periodic report: **1st 2nd X**

Period covered: **from 1.11.2013 to 30.04.2015**

Name, title and organisation of the scientific representative of the project's coordinator[1]:

Dr. Dima Shepelyansky

Directeur de recherche au CNRS

Lab de Phys. Theorique, Universite Paul Sabatier, 31062 Toulouse, France

Tel: +331 5 61556068, Fax: +33 5 61556065, Secr.: +33 5 61557572

E-mail: dima@irsamc.ups-tlse.fr; URL: www.quantware.ups-tlse.fr/dima

Project website address: www.quantware.ups-tlse.fr/FETNADINE/

[1] Usually the contact person of the coordinator as specified in Art. 8.1. of the grant agreement

NADINE DELIVERABLE D2.2.

It is based on milestones M6 (WP2.3), M11 (WP2.4) with deliverable publications:

[12] P1.12 K.M.Frahm, Y.-H.Eom and D.L. Shepelyansky, **"Google matrix of the citation network of Physical Review"**, submitted to Phys. Rev. E Oct 21, 2013 (arXiv:1310.5624 [physics.soc-ph], 2013); published Phys. Rev. E v.89, p.052814 (2014) [M6-WP2.3]

[reported in period 1]

[35] P1.15 L.Ermann, K.M.Frahm and D.L.Shepelyansky, **"Google matrix analysis of directed networks"**, submitted to Rev. Mod. Phys. (2014) (arXiv:1409.0428[physics.soc-ph]) [M6-WP2.3; M11-WP2.4]

[40] P1.20 O.V.Zhirov and D.L.Shepelyansky, **"Anderson transition for Google matrix eigenstates"**, Ann. der Physik (Berlin) DOI 10.1002/andp.201500110 (2015) (arXiv:1501.03371[q-fin.ST]) [M11-WP2.4]

Google matrix of the citation network of Physical Review

Klaus M. Frahm, Young-Ho Eom, and Dima L. Shepelyansky
Laboratoire de Physique Théorique du CNRS, IRSAMC, Université de Toulouse, UPS, 31062 Toulouse, France
(Received 21 October 2013; published 28 May 2014)

We study the statistical properties of spectrum and eigenstates of the Google matrix of the citation network of Physical Review for the period 1893–2009. The main fraction of complex eigenvalues with largest modulus is determined numerically by different methods based on high-precision computations with up to $p = 16\,384$ binary digits that allow us to resolve hard numerical problems for small eigenvalues. The nearly nilpotent matrix structure allows us to obtain a semianalytical computation of eigenvalues. We find that the spectrum is characterized by the fractal Weyl law with a fractal dimension $d_f \approx 1$. It is found that the majority of eigenvectors are located in a localized phase. The statistical distribution of articles in the PageRank-CheiRank plane is established providing a better understanding of information flows on the network. The concept of ImpactRank is proposed to determine an influence domain of a given article. We also discuss the properties of random matrix models of Perron-Frobenius operators.

DOI: 10.1103/PhysRevE.89.052814 PACS number(s): 89.75.Hc, 89.20.Hh, 89.75.Fb

I. INTRODUCTION

The development of the Internet has led to emergence of various types of complex directed networks created by modern society. The size of such networks has grown rapidly going beyond 10 billion in the last two decades for the World Wide Web (WWW). Thus the development of mathematical tools for the statistical analysis of such networks has become of primary importance. In 1998 Brin and Page proposed the analysis of WWW on the basis of the PageRank vector of the associated Google matrix constructed for a directed network [1]. The mathematical foundations of this analysis are based on Markov chains [2] and Perron-Frobenius operators [3]. The PageRank algorithm allows us to compute the ranking of network nodes and is known to be at the heart of modern search engines [4]. However, in many respects the statement of Brin and Page that "Despite the importance of large-scale search engines on the web, very little academic research has been done on them" [1] still remains valid at present. In our opinion, this is related to the fact that the Google matrix G belongs to a new class of operators which had been rarely studied in physical systems. Indeed, the physical systems are usually described by Hermitian or unitary matrices for which random matrix theory [5] captures many universal properties. In contrast, the Perron-Frobenium operators and Google matrix have eigenvalues distributed in the complex plane belonging to another class of operators.

The Google matrix is constructed from the adjacency matrix A_{ij}, which has unit elements if there is a link pointing from node j to node i and zero otherwise. Then the matrix of Markov transitions is constructed by normalizing elements of each column to unity ($S_{ij} = A_{ij}/\sum_i A_{ij}$, $\sum_j S_{ij} = 1$) and replacing columns with only zero elements (*dangling nodes*) by $1/N$, with N being the matrix size. After that the Google matrix of the network takes the form [1,4]

$$G_{ij} = \alpha S_{ij} + (1-\alpha)/N. \quad (1)$$

The damping parameter α in the WWW context describes the probability $(1-\alpha)$ to jump to any node for a random surfer. For WWW the Google search engine uses $\alpha \approx 0.85$ [4]. The PageRank vector P_i is the right eigenvector of G at $\lambda = 1$ ($\alpha < 1$). According to the Perron-Frobenius theorem [3], P_i components are positive and represent the probability to find a random surfer on a given node i (in the stationary limit) [4]. All nodes can be ordered in a decreasing order of probability $P(K_i)$ with highest probability at top values of PageRank index $K_i = 1,2,\ldots$.

The distribution of eigenvalues of G can be rather nontrivial with the appearance of the fractal Weyl law and other unusual properties (see, e.g., Refs. [6,7]). For example, a matrix G with random positive matrix elements, normalized to unity in each column, has $N-1$ eigenvalues λ concentrated in a small radius $|\lambda| < 1/\sqrt{3N}$ and one eigenvalue $\lambda = 1$ (see Sec. VII). Such a distribution is drastically different from the eigenvalue distributions found for directed networks with algebraic distribution of links [8] or those found numerically for other directed networks including universities' WWW [9,10], Linux Kernel and Twitter networks [11,12], and Wikipedia networks [13,14]. In fact, even the Albert-Barabási model of preferential attachment [15] still generates the complex spectrum of λ with a large gap ($|\lambda| < 1/2$) [8] being very different from the gapless and strongly degenerate G spectrum of British universities' WWW [10] and Wikipedia [13,14]. Thus it is useful to get a deeper understanding of the spectral properties of directed networks and to develop more advanced models of complex networks which have a spectrum similar to such networks as those of British universities and Wikipedia.

With the aim to understand the spectral properties of a Google matrix of directed networks we study here the citation network of Physical Review (CNPR) for the whole period up to 2009 [16]. This network has $N = 463\,348$ nodes (articles) and $N_\ell = 4\,691\,015$ links. Its network structure is very similar to the tree network since the citations are time ordered (with only a few exceptions of mutual citations of simultaneously published articles). As a result we succeed in developing powerful tools, which allowed us to obtain the spectrum of G in semianalytical way. These results are compared with the spectrum obtained numerically with the help of the powerful Arnoldi method (see its description in Refs. [17,18]). Thus we are able to get a better understanding of the spectral properties of this network. Due to time ordering of article citations there are strong similarities between the CNPR and the network of integers studied recently in Refs. [19].

We note that the PageRank analysis of the CNPR had been performed in Refs. [20–23] showing its efficiency in determining the influential articles of Physical Review. The citation networks are rather generic (see, e.g., Ref. [24]), and hence the extension of PageRank analysis of such networks is an interesting and important task. Here we put the main accent on the spectrum and eigenstates properties of the Google matrix of the CNPR, but we also discuss the properties of two-dimensional (2D) ranking on the PageRank-CheiRank plane developed recently in Refs. [25–27]. We also analyze the properties of ImpactRank, which shows a domain of influence of a given article.

In addition to the whole CNPR we also consider the CNPR without Revview of Modern Physics articles, which has $N = 460\,422$, $N_\ell = 4\,497\,707$. If in the whole CNPR we eliminate future citations (see description below), then this triangular CNPR has $N = 463\,348$, $N_\ell = 4\,684\,496$. Thus on average we have approximately 10 links per node. The network includes all articles of Physical Review from its foundation in 1893 till the end of 2009.

The paper is composed as follows: in Sec. II we present a detailed analysis of the Google matrix spectrum of CNPR, the fractal Weyl law is discussed in Sec. III, properties of eigenstates are discussed in Sec. IV, CheiRank versus PageRank distributions are considered in Sec. V, properties of impact propagation through the network are studied in Sec. VI, certain random matrix models of Google matrix are studied in Sec. VII, and discussion of the results is given in Section VIII.

II. EIGENVALUE SPECTRUM

The Google matrix of CNPR is constructed on the basis of Eq. (1) using citation links from one article to another (see also Refs. [21–23]). The matrix structure for different order representations of articles is shown in Fig. 1. In the top left panel all articles are ordered by time, which generates an almost perfect triangular structure corresponding to time ordering of citations. Still there are a few cases with joint citations of articles which appear almost at the same time. Also there are dangling nodes which generate transitions to all articles with elements $1/N$ in G. This breaks the triangular structure, and as we will see later it is just the combination of these dangling node contributions with the other nonvanishing matrix elements [see also Eq. (4)] which will allow us to formulate a semianalytical theory to determine the eigenvalue spectrum.

The triangular matrix structure is also well visible in the middle left panel where articles are time ordered within each Physical Review journal. The left bottom panel shows the matrix elements for each Physical Review journal when inside each journal the articles are ordered by their PageRank index K. The right panels show the matrix elements of G on different scales, when all articles are ordered by the PageRank index K. The top two right panels have a relatively small number of nonzero matrix elements showing that the top PageRank articles rarely quote other top PageRank articles.

The dependence of the number of no-zero links N_G, between nodes with PageRank index being less than K, on K is shown in Fig. 2 (left panel). We see that compared to the other networks of universities, Wikipedia, and Twitter studied in Ref. [13] we have for CNPR the lowest values of N_G/K practically for all available K values. This reflects weak links between top PageRank articles of CNPR in contrast with Twitter, which has a very high interconnection between top PageRank nodes. Since the matrix elements $G_{KK'}$ are inversely proportional to the number of links, we have very strong average matrix elements for CNPR at top K values (see Fig. 2, right panel).

In the following we present the results of numerical and analytical analysis of the spectrum of the CNPR matrix G.

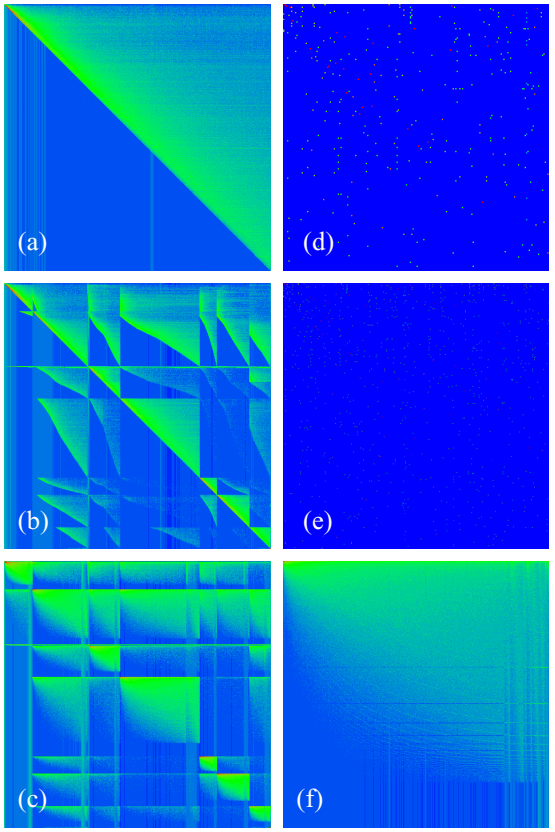

FIG. 1. (Color online) Different order representations of the Google matrix of the CNPR ($\alpha = 1$). The left panels show (a) the density of matrix elements $G_{tt'}$ in the basis of the publication time index t (and t'); (b) the density of matrix elements in the basis of journal ordering according to Phys. Rev. Series I, Phys. Rev., Phys. Rev. Lett., Rev. Mod. Phys., Phys. Rev. A, B, C, D, E, Phys. Rev. STAB, and Phys. Rev. STPER with time ordering inside each journal; (c) the same as (b) but with PageRank index ordering inside each journal. Note that the journals Phys. Rev. Series I, Phys. Rev. STAB, and Phys. Rev. STPER are not clearly visible due to a small number of published papers. Also Rev. Mod. Phys. appears only as a thin line with 2–3 pixels (out of 500) due to a limited number of published papers. The panels (a), (b), (c), and (f) show the coarse-grained density of matrix elements done on 500×500 square cells for the entire network. In panels (d), (e), and (f) the matrix elements $G_{KK'}$ are shown in the basis of PageRank index K (and K') with the range $1 \leqslant K,K' \leqslant 200$ (d); $1 \leqslant K,K' \leqslant 400$ (e); $1 \leqslant K,K' \leqslant N$ (f). Color shows the amplitude (or density) of matrix elements G changing from blue/black for zero value to red/gray at maximum value. The PageRank index K is determined from the PageRank vector at $\alpha = 0.85$.

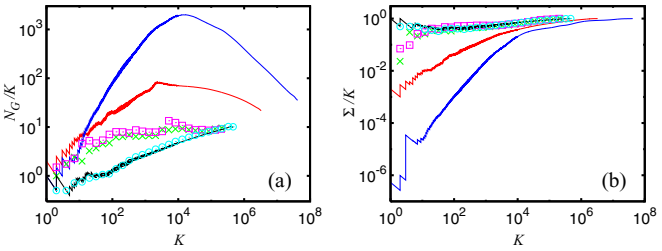

FIG. 2. (Color online) (a) Dependence of the linear density N_G/K of nonzero elements of the adjacency matrix among top PageRank nodes on the PageRank index K for the networks of Twitter (top blue/black curve), Wikipedia (second from top red/gray curve), Oxford University 2006 (magenta/gray boxes), and Cambridge University 2006 (green/gray crosses), with data taken from Ref. [12], and Physical Review all journals (cyan/gray circles) and Physical Review without Rep. Mod. Phys. (bottom black curve). (b) Dependence of the quantity Σ/K on the PageRank index K with $\Sigma = \sum_{K_1<K,\,K_2<K} G_{K_1,K_2}$ being the weight of the Google matrix elements inside the $K \times K$ square of top PageRank indexes. The curves correspond to the same networks as in (a): Physical Review without Rep. Mod. Phys. (top black curve), Physical Review all journals (cyan/gray circles), Oxford University 2006 (magenta/gray boxes), Cambridge University 2006 (green/gray crosses), Wikipedia (second bottom red/gray curve), and Twitter (bottom blue/black curve).

A. Nearly nilpotent matrix structure

The triangular structure of the CNPR Google matrix in the time index (see Fig. 1) has important consequences for the eigenvalue spectrum λ defined by the equation for the eigenstates $\psi_i(j)$:

$$\sum_{j'} G_{jj'}\psi_i(j') = \lambda_i \psi_i(j). \tag{2}$$

The spectrum of G at $\alpha = 1$, or the spectrum of S, obtained by the Arnoldi method [17,18] with the Arnoldi dimension $n_A = 8000$, is shown in Fig. 3. For comparison we also show the case of a reduced CNPR without Review of Modern Physics. We see that the spectrum of the reduced case is rather similar to the spectrum of the full CNPR.

The nodes can be decomposed in invariant subspace nodes and core space nodes, and the matrix S can be written in the block structure [10]:

$$S = \begin{pmatrix} S_{ss} & S_{sc} \\ 0 & S_{cc} \end{pmatrix}, \tag{3}$$

where S_{ss} contains the links from subspace nodes to other subspace nodes, S_{cc} the links from core space nodes to core space nodes, and S_{sc} some coupling links from the core space to the invariant subspaces. The subspace-subspace block S_{ss} is actually composed of (potentially) many diagonal blocks for each of the invariant subspaces. Each of these blocks corresponds to a column sum-normalized matrix of the same type as G and has therefore at least one unit eigenvalue, thus explaining a possible high degeneracy of the eigenvalue $\lambda = 1$ of S. This structure is discussed in detail in Ref. [10]. The university networks discussed in Ref. [10] had a considerable number of subspace nodes (about 20%) with a high degeneracy

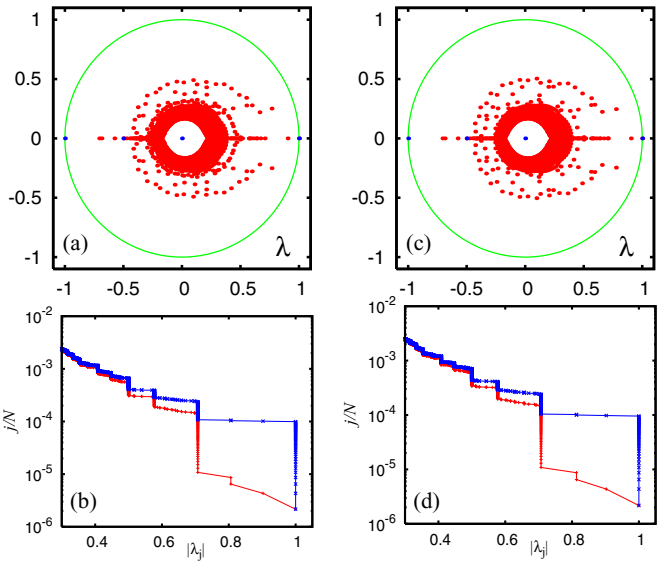

FIG. 3. (Color online) Spectrum of S for CNPR (reduced CNPR without Rev. Mod. Phys.) shown on left panels (right panels). (a, c) Subspace eigenvalues (blue/black dots) and core space eigenvalues (red/gray dots) in the λ plane (green/gray curve shows unit circle); there are 27 (26) invariant subspaces, with maximal dimension 6 (6), and the sum of all subspace dimensions is $N_s = 71$ (75). The core space eigenvalues are obtained from the Arnoldi method applied to the core space subblock S_{cc} of S with Arnoldi dimension $n_A = 8000$ as explained in Ref. [10] and using standard double-precision arithmetic. (b, d) Fraction j/N of eigenvalues, shown in a logarithmic scale, with $|\lambda| > |\lambda_j|$ for the core space eigenvalues (red/gray bottom curve) and all eigenvalues (blue/black top curve) from raw data of top panels. The number of eigenvalues with $|\lambda_j| = 1$ is 45 (43) of which 27 (26) are at $\lambda_j = 1$; this number is identical to the number of invariant subspaces which have each one unit eigenvalue.

$\sim 10^3$ of the leading unit eigenvalue. However, for the CNPR the number of subspace nodes and unit eigenvalues is quite small (see the figure caption of Fig. 3 for detailed values).

A network with a similar triangular structure, constructed from factor decompositions of integer numbers, was previously studied in Ref. [19]. There it was analytically shown that the corresponding matrix S has only a small number of nonvanishing eigenvalues and that the numerical diagonalization methods, including the Arnoldi method, are facing subtle difficulties of numerical stability due to large Jordan blocks associated to the highly degenerate zero eigenvalue. The numerical diagonalization of these Jordan blocks is highly sensitive to numerical round-off errors. For example, a perturbed Jordan block of dimension D associated to the eigenvalue zero and with a perturbation ε in the opposite corner has eigenvalues on a complex circle of radius $\varepsilon^{1/D}$ [19], which may became quite large for sufficient large D even for $\varepsilon \sim 10^{-15}$. Therefore in the presence of many such Jordan blocks the numerical diagonalization methods create rather big "artificial clouds" of incorrect eigenvalues.

In the examples studied in Ref. [19] these clouds extended up to eigenvalues $|\lambda| \approx 0.01$. The spectrum for the Physical Review network shown in Fig. 3 shows also a sudden

increase of the density of eigenvalues below $|\lambda| \approx 0.3$–0.4, and one needs to be concerned if these eigenvalues are numerically correct or only an artifact of the same type of numerical instability. Actually there is a quite simple way to verify that they are not reliable due to problems in the numerical evaluation. For this we apply to the network or the numerical algorithm (in the computer program) certain transformations or modifications which are *mathematically neutral* or *equivalent*, e.g., a permutation of the index numbers of the network nodes but keeping the same network-link structure, or simply changing the evaluation order in the sums used for the scalar products between vectors (in the Gram-Schmidt orthogonalization for the Arnoldi method). All these modifications should in theory not modify the results (assuming that all computations could be done with *infinite precision*), but in numerical computations on a computer with finite precision they modify the round-off errors. It turns out indeed that the modifications of the initially small round-off errors induce very strong, completely random modifications, for all eigenvalues below $|\lambda| \approx 0.3$–0.4, clearly indicating that the latter are numerically not accurate. Apparently the problematic numerical eigenvalue errors due to large Jordan blocks $\sim \varepsilon^{1/D}$ with $D \sim 10^2$ is quite stronger in the Physical Review citation network than in the previously studied integer network [19].

The theory of Ref. [19] is based on the exact triangular structure of the matrix S_0, which appears in the representation of $S = S_0 + ed^T/N$ [see also Eq. (4)]. In fact, the matrix S_0 is obtained from the adjacency matrix by normalizing the sum of the elements in nonvanishing columns to unity and simply keeping at zero vanishing columns. For the network of integers [19] this matrix is nilpotent with $S_0^l = 0$ for a certain modest value of l being much smaller than the network size $l \ll N$. The nilpotency is very relevant in the paper for two reasons: first, it is responsible for the numerical problems to compute the eigenvalues by standard methods (see the next point), and, second, it is also partly the solution by allowing a semianalytical approach to determine the eigenvalues in a different way.

For CNPR the matrix S_0 is not exactly nilpotent despite the overall triangular matrix structure visible in Fig. 1. Even though most of the nonvanishing matrix elements $(S_0)_{tt'}$ (whose total number is equal to the number of links $N_\ell = 4\,691\,015$) are in the upper triangle $t < t'$ there are a few nonvanishing elements in the lower triangle $t > t'$ (whose number is 12 126 corresponding to 0.26% of the total number of links [28]). The reason is that in most cases papers cite other papers published earlier but in certain situations for papers with a close publication date the citation order does not always coincide with the publication order. In some cases two papers even mutually cite each other. In the following we will call these cases "future citations." The rare nonvanishing matrix elements due to future citations are not visible in the coarse-grained matrix representation of Fig. 1, but they are responsible for the fact that S_0 of CMPR is not nilpotent and that there are also a few invariant subspaces. On a purely triangular network one can easily show the absence of invariant subspaces (smaller than the full network size) when taking into account the extra columns due to the dangling nodes.

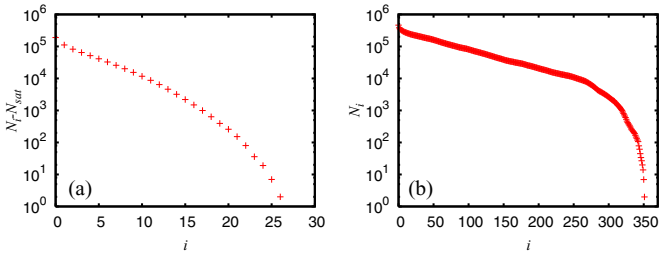

FIG. 4. (Color online) Number of occupied nodes N_i (i.e., positive elements) in the vector $S_0^i e$ versus iteration number i (red/gray crosses) for the CNPR (a) and the triangular CNPR (b). In both cases the initial value is the network size $N_0 = N = 463\,348$. For the CNPR N_i saturates at $N_i = N_{\rm sat} = 273\,490 \approx 0.590N$ for $i \geqslant 27$ while for the triangular CNPR N_i saturates at $N_i = 0$ for $i \geqslant 352$ confirming the nilpotent structure of S_0. In (a) the quantity $N_i - N_{\rm sat}$ is shown in order to increase visibility in the logarithmic scale.

However, despite the effect of the future citations the matrix S_0 is still partly nilpotent. This can be seen by multiplying a uniform initial vector e (with all components being 1) by the matrix S_0 and counting after each iteration the number N_i of nonvanishing entries [29] in the resulting vector $S_0^i e$. For a nilpotent matrix S_0 with $S_0^l = 0$ the number N_i becomes obviously zero for $i \geqslant l$. On the other hand, since the components of e and the nonvanishing matrix elements of S_0 are positive, one can easily verify that the condition $S_0^l e = 0$ for some value l also implies $S_0^l \psi = 0$ for an arbitrary initial (even complex) vector ψ, which shows that S_0 must be nilpotent with $S_0^l = 0$.

In Fig. 4 we see that for the CNPR the value of N_i saturates at a value $N_{\rm sat} = 273\,490$ for $i \geqslant 27$, which is 59% of the total number of nodes $N = 463\,348$ in the network. On one hand the (small) number of future citations ensures that the saturation value of N_i is not zero, but on the other hand it is smaller than the total number of nodes by a macroscopic factor. Mathematically the first iteration $e \to S_0 e$ removes the nodes corresponding to empty (vanishing) lines of the matrix S_0, and the next iterations remove the nodes whose lines in S_0 have become empty after having removed from the network the nonoccupied nodes due to previous iterations. For each node removed during this iteration process one can construct a vector belonging to the Jordan subspace of S_0 associated to the eigenvalue 0. In the following we call this subspace *generalized kernel*. It contains all eigenvectors of S_0^j associated to the eigenvalue 0 where the integer j is the size of the largest 0-eigenvalue Jordan block. Obviously the dimension of this generalized kernel of S_0 is larger or equal than $N - N_{\rm sat} = 189\,857$, but we will see later that its actual dimension is even larger and quite close to N. We will argue below that most (but not all) of the vectors in the generalized kernel of S_0 also belong to the generalized kernel of S, which differs from S_0 by the extra contributions due to the dangling nodes. The high dimension of the generalized kernel containing many large 0-eigenvalue Jordan subspaces explains very clearly the numerical problem due to which the eigenvalues obtained by the double-precision Arnoldi method are not reliable for $|\lambda| <$ 0.3–0.4.

B. Spectrum for the triangular CNPR

In order to extend the theory for the triangular matrices developed in Ref. [19] we consider the triangular CNPR obtained by removing all future citation links $t' \to t$ with $t \geqslant t'$ from the original CNPR. The resulting matrix S_0 of this reduced network is now indeed nilpotent with $S_0^{l-1} \neq 0$, $S_0^l = 0$, and $l = 352$, which is much smaller than the network size. This is clearly seen from Fig. 4, showing that N_i, calculated from the triangular CNPR, indeed saturates at $N_i = 0$ for $i \geqslant 352$. According to the arguments of Ref. [19], and additional demonstrations given below, there are at most only $l = 352$ nonzero eigenvalues of the Google matrix at $\alpha = 1$. This matrix has the form

$$S = S_0 + (1/N)\, e\, d^T, \tag{4}$$

where d and e are two vectors with $e(n) = 1$ for all nodes $n = 1, \ldots, N$ and $d(n) = 1$ for dangling nodes n (corresponding to vanishing columns in S_0) and $d(n) = 0$ for the other nodes. In the following we call d the dangling vector. The extra contribution $e\, d^T/N$ just replaces the empty columns (of S_0) with $1/N$ entries at each element and d^T is the line vector obtained as the transpose of the column vector d. In Appendix A we extend the approach of Ref. [19] showing analytically that the matrix S has exactly $l = 352 \ll N$ nonvanishing eigenvalues, which are given as the zeros of the reduced polynomial given in Eq. (A4), and that it is possible to define a closed representation space for the matrix S of dimension l leading to an $l \times l$ representation matrix $\bar{S}$ given by Eq. (A8) whose eigenvalues are exactly the zeros of the reduced polynomial.

In the left panel of Fig. 5 we compare the core space spectrum of S for CNPR and triangular CNPR (data are obtained by the Arnoldi method with $n_A = 4000$ and standard double precision). We see that the largest complex eigenvalues are rather close for both cases, but in the full network we have many eigenvalues on the real axis (with $\lambda < -0.3$ or $\lambda > 0.4$), which are absent for the triangular CNPR. Furthermore, both cases suffer from the same problem of numerical instability due to large Jordan blocks.

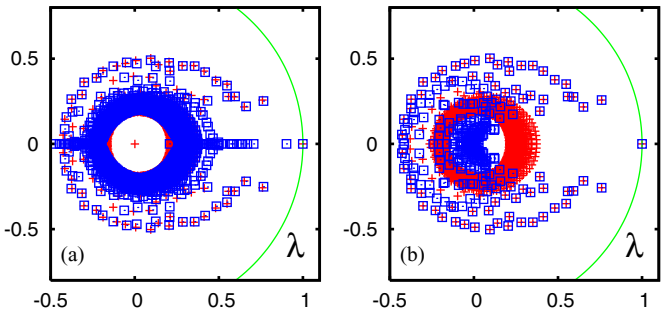

FIG. 5. (Color online) (a) Comparison of the core space eigenvalue spectrum of S for CNPR (blue/black squares) and triangular CNPR (red/gray crosses). Both spectra are calculated by the Arnoldi method with $n_A = 4000$ and standard double precision. (b) Comparison of the numerically determined nonvanishing 352 eigenvalues obtained from the representation matrix (A8) (blue/black squares) with the spectrum of triangular CNPR (red/gray crosses) already shown in the left panel. Numerics is done with standard double precision.

In the right panel of Fig. 5 we compare the numerical double-precision spectra of the representation matrix $\bar{S}$ with the results of the Arnoldi method with double precision and the uniform initial vector e as start vector for the Arnoldi iterations (applied to the triangular CNPR). In Appendix B we explain that the Arnoldi method with this initial vector should in theory (in absence of rounding errors) also exactly provide the l eigenvalues of $\bar{S}$ since by construction it explores the same l-dimensional S-invariant representation space that was used for the construction of $\bar{S}$ (in Appendix A). The fact that both spectra of the right panel of Fig. 5 differ is therefore a clear effect of numerical errors, and actually both cases suffer from different numerical problems (see Appendix B for details). A different, and in principle highly efficient, computational method is to calculate the spectrum of the triangular CNPR by determining numerically the l zeros of the reduced polynomial (A4), but according to the further discussion in Appendix B there are also numerical problems for this. Actually this method requires the help the GNU Multiple Precision Arithmetic Library (GMP library) [30] using 256 binary digits. Also the Arnoldi method can be improved by GMP library (see Appendix B for details) even though this is quite expensive in computational time and memory usage but still feasible (using up to 1280 binary digits). Below we will also present results (for the spectrum of the full CNPR) based on a new method using the GMP library with up to 16 384 binary digits.

In Fig. 6 we compare the exact spectrum of the triangular CNPR obtained by the high-precision determination of the zeros of the reduced polynomial (using 256 binary digits) with the spectra of the Arnold method for 52 binary digits (corresponding to the mantissa of double-precision numbers), 256, 512, and 1280 binary digits. Here we use for the Arnoldi method a uniform initial vector and the Arnold dimension $n_A = l = 352$. In this case, as explained in Appendix B, in theory the Arnoldi method should provide the exact $l = 352$ nonvanishing eigenvalues (in the absence of round-off errors).

However, with the precision of 52 bits we have a considerable number of eigenvalues on a circle of radius ≈ 0.3 centered at 0.05, indicating a strong influence of round-off errors due to the Jordan blocks. Increasing the precision to 256 (or 512) binary digits implies that the number of correct eigenvalue increases and the radius of this circle decreases to 0.13 (or 0.1), and in particular it does not extend to all angles. We have to increase the precision of the Arnoldi method to 1280 binary digits to have a perfect numerical confirmation that the Arnoldi method explores the exact invariant subspace of dimension $l = 352$ and generated by the vectors v_j (see Appendix A). In this case the eigenvalues obtained from the Arnoldi method and the high-precision zeros of the reduced polynomial coincide with an error below 10^{-14} and in particular the Arnoldi method provide a nearly vanishing coupling matrix element at the last iteration, confirming that there is indeed an exact decoupling of the Arnoldi matrix and an invariant closed subspace of dimension 352.

The results shown in Fig. 6 clearly confirm the above theory and the scenario of the strong influence of Jordan blocks on the round-off errors. In particular, we find that in order to increase

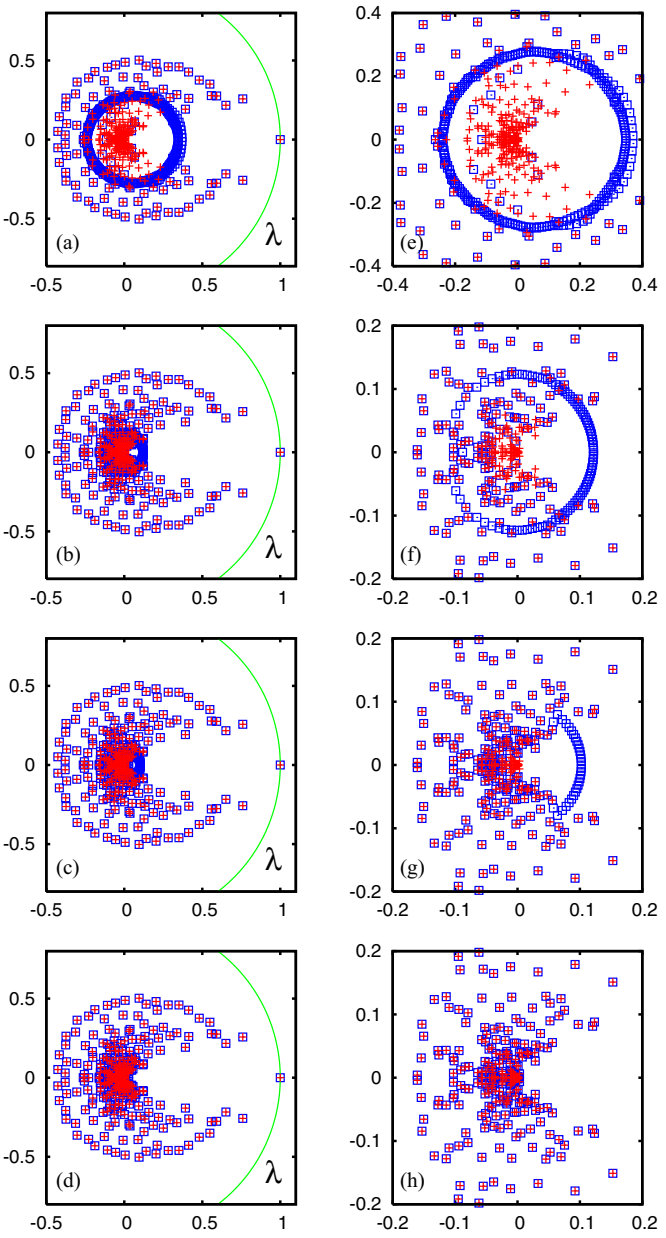

FIG. 6. (Color online) Comparison of the numerically accurate 352 nonvanishing eigenvalues of S matrix of triangular CNPR, determined by the Newton-Maehly method applied to the reduced polynomial (A4) with a high-precision calculation of 256 binary digits (red/gray crosses, all panels), with eigenvalues obtained by the Arnoldi method at different numerical precisions (for the determination of the Arnoldi matrix) for triangular CNPR and Arnoldi dimension $n_A = 352$ (blue/black squares, all panels). The first row corresponds to the numerical precision of 52 binary digits for standard double-precision arithmetic. The second (third, fourth) row corresponds to the precision of 256 (512, 1280) binary digits. All high-precision calculations are done with the library GMP [30]. The panels in the left column show the complete spectra and the panels in the right columns show the spectra in a zoomed range: $-0.4 \leqslant \mathrm{Re}(\lambda),\mathrm{Im}(\lambda) \leqslant < 0.4$ for the first row or $-0.2 \leqslant \mathrm{Re}(\lambda),\mathrm{Im}(\lambda) \leqslant 0.2$ for the second, third, and fourth rows.

the numerical precision it is only necessary to implement the first step of the method, the Arnoldi iteration, using high-precision numbers while the numerical diagonalization of the Arnoldi representation matrix can still be done using standard double-precision arithmetic. We also observe that even for the case with lowest precision of 52 binary digits the eigenvalues obtained by the Arnoldi method are numerically accurate provided that there are well outside the circle (or cloud) of numerically incorrect eigenvalues.

C. High-precision spectrum of the whole CNPR

Based on the observation that a high-precision implementation of the Arnoldi method is useful for the triangular CNPR, we now apply the high-precision Arnoldi method with 256, 512, and 756 binary digits and $n_A = 2000$ to the original CNPR. The results for the core space eigenvalues are shown in Fig. 7, where we compare the spectrum of the highest precision of 756 binary digits with lower-precision spectra of 52, 256, and 512 binary digits. As in Fig. 6 for the triangular CNPR, for CNPR we also observe that the radius and angular extension of the cloud or circle of incorrect Jordan block eigenvalues decrease with increasing precision. Despite the lower number of $n_A = 2000$ as compared to $n_A = 8000$ of Fig. 3 the number of accurate eigenvalues with 756-bit precision is certainly considerably higher.

The higher-precision Arnoldi method certainly improves the quality of the smaller eigenvalues, e.g., for $|\lambda| < 0.3 - 0.4$,

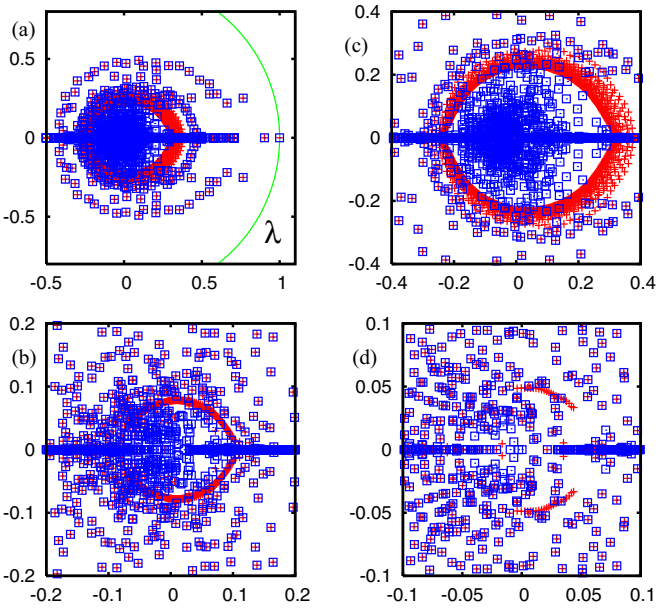

FIG. 7. (Color online) Comparison of the core space eigenvalue spectrum of S of CNPR, obtained by the high-precision Arnoldi method using 768 binary digits (blue/black squares, all panels), with lower-precision data of the Arnoldi method (red/gray crosses). In both top panels the red/gray crosses correspond to double precision with 52 binary digits [extended range in (a) and zoomed range in (c)]. In the bottom (b) [or (d)] panel red/gray crosses correspond to the numerical precision of 256 (or 512) binary digits. In these two cases only a zoomed range is shown. The eigenvalues outside the zoomed ranges coincide for both data sets up to graphical precision. In all cases the Arnoldi dimension is $n_A = 2000$. High-precision calculations are done with the library GMP [30].

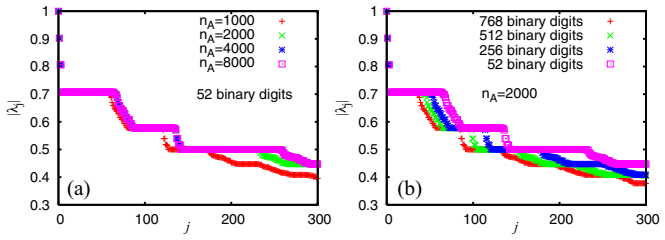

FIG. 8. (Color online) Modulus $|\lambda_j|$ of the core space eigenvalues of S of CNPR, obtained by the Arnoldi method, shown versus level number j. (a) Data for standard double precision with 52 binary digits with different Arnoldi dimensions $1000 \leqslant n_A \leqslant 8000$. (b) Data for Arnoldi dimension $n_A = 2000$ with different numerical precisions between 52 and 768 binary digits.

but it also implies a strange shortcoming as far as the degeneracies of certain particular eigenvalues are concerned. This can be seen in Fig. 8, which shows the core space eigenvalues $|\lambda_j|$ versus the level number j for various values of the Arnoldi dimension and the precision. In these curves we observe flat plateaux at certain values $|\lambda_j| = 1/\sqrt{n}$ with $n = 2, 3, 4, 5, \ldots$ corresponding to degenerate eigenvalues which turn out to be real but with positive or negative values: $\lambda_j = \pm 1/\sqrt{n}$. For fixed standard double-precision arithmetic with 52 binary digits the degeneracies increase with increasing Arnoldi dimension and seem to saturate for $n_A \geqslant 4000$. However, at the given value of $n_A = 2000$ the degeneracies *decrease* with increasing precision of the Arnoldi method. Apparently the higher-precision Arnoldi method is less able to determine the correct degeneracy of a degenerate eigenvalue.

This point can be understood as follows. In theory, assuming perfect precision, the simple version of the Arnoldi method used here (in contrast to more complicated block Arnoldi methods) can determine only one eigenvector for a degenerate eigenvalue. The reason is that for a degenerate eigenvalue we have a particular linear combination of the eigenvectors for this eigenvalue which contribute in any initial vector (in other words "one particular" eigenvector for this eigenvalue), and during the Arnoldi iteration this particular eigenvector will be perfectly conserved and the generated Krylov space will contain only this and no other eigenvector for this eigenvalue. However, due to round-off errors we obtain at each step new random contributions from other eigenvectors of the same eigenvalue, and it is due only to these round-off errors that we can see the flat plateaux in Fig. 8. Obviously, increasing the precision reduces this round-off error effect, and the flat plateaux are indeed considerably smaller for higher precisions.

The question arises about the origin of the degenerate eigenvalues in the core space spectrum. In other examples, such as the WWW for certain university networks [10], the degeneracies, especially of the leading eigenvalue 1, could be treated by separating and diagonalizing the exact subspaces, and the remaining core space spectrum contained much less or nearly no degenerate eigenvalues. However, here for the CNPR we have "only" 27 subspaces with maximal dimension of 6 containing 71 nodes in total. The eigenvalues due to these subspaces are 1, -1, -0.5, and 0 with degeneracies 27, 18, 4, and 22 (see blue dots in the upper panels of Fig. 3). These exact subspaces exist due only to the modest number of future citation links. Even when we take care that in all cases the Arnoldi method is applied to the core space without these 71 subspace nodes, there still remain many degenerate eigenvalues in the core space spectrum.

In Appendix C we explain how the degenerated core space eigenvalues of S can be obtained as degenerate subspace eigenvalues of S_0 (i.e., neglecting the dangling node contributions when determining the invariant subspaces). To be precise it turns out that the core space eigenvalues of S are decomposed in two groups: the first group is related to degenerate subspace eigenvalues of S_0 and can be determined by a scheme described in Appendix C, and the second group of eigenvalues is given as zeros of a certain rational function ((D1),) which can be evaluated by the series (D3) which converges only for $|\lambda| > \rho_1$ with $\rho_1 \approx 0.902$. To determine the zeros of the rational function, outside the range of convergence, one can employ an argument of analytical continuation using a new method, called "rational interpolation method" described in detail in Appendix D. Without going into much detail here, we mention that the main idea of this method is to evaluate this rational function at many support points on the complex unit circle where the series (D3) converges well and then to use these values to interpolate the rational function (D1) by a simpler rational function for which the zeros can be determined numerically well even if they are inside the unit circle (where the initial series does not converge). For this scheme it is also very important to use high-precision computations. Typically for a given precision of p binary digits one may chose a certain number n_R of eigenvalues to be determined choosing the appropriate number of support points (either $2n_R + 1$ or $2n_R + 2$ depending on the variant of the method; see also Appendix D). Provided that n_R is not neither too small nor too large (depending on the value of p) one obtains very reliable core space eigenvalues of S of the second group.

For example, as can be seen in Fig. 9, for $p = 1024$ we obtain $n_R = 300$ eigenvalues for which the big majority coincides numerically (error $\sim 10^{-14}$) with the eigenvalues obtained from the high-precision Arnoldi method for 768 binary digits, and furthermore both variants of the rational interpolation method provide identical spectra.

However, for $n_R = 340$ some of the zeros do not coincide with eigenvalues of S, and most of these deviating zeros lie close to the unit circle. We can even somehow distinguish between "good" zeros (associated to eigenvalues of S) being identical for both variants of the method and "bad" artificial zeros, which are completely different for both variants (see Fig. 9). We note that for the case of too large n_R values the artificial zeros are extremely sensitive to numerical round-off errors (in the high-precision variables) and that they change strongly, when slightly modifying the support points (e.g., a random modification $\sim 10^{-18}$ or simply changing their order in the interpolation scheme) or when changing the precise numerical algorithm (e.g., between a direct sum or Horner scheme for the evaluation of the series of the rational function). Furthermore, they do not respect the symmetry that the zeros should come in pairs of complex conjugate numbers in case of complex zeros. This is because Thiele's rational interpolation scheme breaks the symmetry due to complex conjugation once round-off errors become relevant.

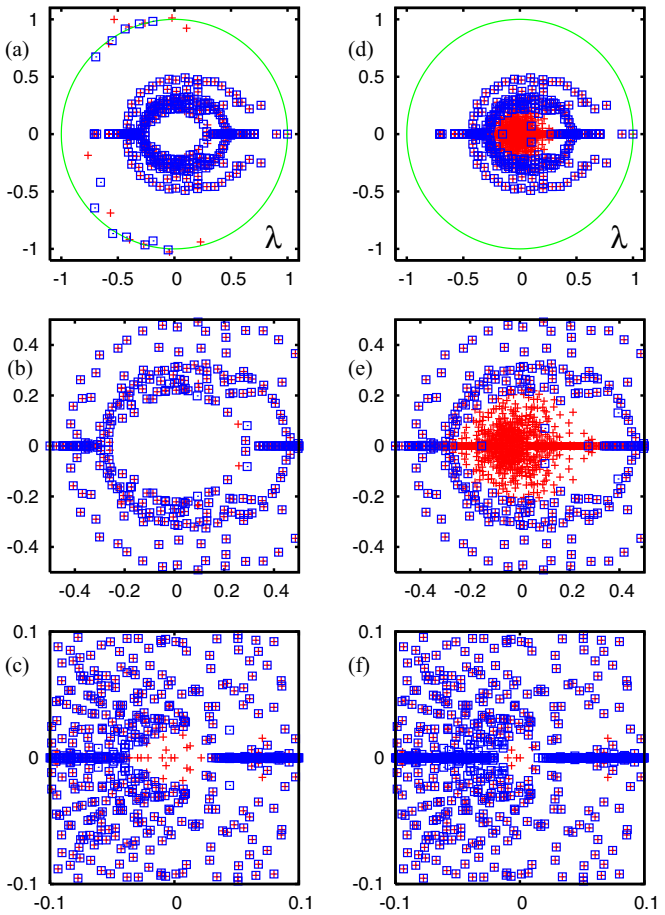

FIG. 9. (Color online) (a) Comparison of $n_R = 340$ core space eigenvalues of S for CNPR obtained by two variants of the rational interpolation method (see text) with the numerical precision of $p = 1024$ binary digits, 681 support points (first variant, red/gray crosses), or 682 support points (second variant, blue/black squares). (d) Comparison of the core space eigenvalues of CNPR obtained by the high-precision Arnoldi method with $n_A = 2000$ and $p = 768$ binary digits (red/gray crosses, same data as blue/black squares in Fig. 7) with the eigenvalues obtained by (both variants of) the rational interpolation method with the numerical precision of $p = 1024$ binary digits and $n_R = 300$ eigenvalues (blue/black squares). Here both variants with 601 or 602 support points provide identical spectra (differences below 10^{-14}). (b, e) Same as panels (a) and (d) with a zoomed range: $-0.5 \leqslant \mathrm{Re}(\lambda), \mathrm{Im}(\lambda) \leqslant 0.5$. (c) Comparison of the core space spectra obtained by the high-precision Arnoldi method (red/gray crosses, $n_A = 2000$ and $p = 768$) and by the rational interpolation method with $p = 12\,288$, $n_R = 2000$ eigenvalues (blue/black squares). (f) Same as (c) with $p = 16\,384$, $n_R = 2500$ for the rational interpolation method. Both panels (c) and (f) are shown in a zoomed range: $-0.1 \leqslant \mathrm{Re}(\lambda), \mathrm{Im}(\lambda) \leqslant 0.1$. Eigenvalues outside the shown range coincide up to graphical precision, and both variants of the rational interpolation method provide numerically identical spectra.

However, we have carefully verified that for the proper values of n_R not being too large (e.g., $n_R = 300$ for $p = 1024$) the obtained zeros are numerically identical (with 52 binary digits in the final result) with respect to small changes of the support points (or their order) or with respect to different numerical algorithms and that they respect perfectly the symmetry due to complex conjugation.

This method, despite the necessity of high-precision calculations, is not very expensive, especially for the memory usage, if compared with the high-precision Arnoldi method. Furthermore, its efficiency for the computation time can be improved by the trick of summing up the largest terms in the series (D3) as a geometrical series which allows to reduce the cutoff value of l by a good factor 3, i.e., replacing $\rho_1 \approx 0.902$ by $\rho_2 = 1/\sqrt{2} \approx 0.707$ in the estimate (D4) of l which gives $l \approx 2\,p + \mathrm{const}$. We have increased the number of binary digits up to $p = 16\,384$, and we find that for $p = 1024, 2048, 4096, 6144, 8192, 12\,288, 16\,384$ we may use $n_R = 300, 500, 900, 1200, 1500, 2000, 2500$ and still avoid the appearance of artificial zeros. In Fig. 9 we also compare the result of the highest precisions $p = 12\,288$ (and $p = 16\,384$) using $n_R = 2000$ ($n_R = 2500$) with the high-precision Arnoldi method with $n_A = 2000$ and $p = 768$, and these spectra coincide well apart from a minor number of smallest eigenvalues. In general, the complex isolated eigenvalues converge very well (with increasing values of p and n_R), while the strongly clustered eigenvalues on the real axis have more difficulties to converge. Comparing the results between $n_R = 2000$ and $n_R = 2500$ we see that the complex eigenvalues coincide on graphical precision for $|\lambda| \geqslant 0.04$ and the real eigenvalues for $|\lambda| \geqslant 0.1$. The Arnoldi method has even more difficulties on the real axis (convergence roughly for $|\lambda| \geqslant 0.15$) since it implicitly has to take care of the highly degenerate eigenvalues of the first group and for which it has difficulties to correctly find the degeneracies (see also Fig. 8).

In Fig. 10 we show as a summary the highest-precision spectra of S with core space eigenvalues obtained by the Arnoldi method or the rational interpolation method (both at best parameter choices) and taking into account the direct subspace eigenvalues of S and the above determined eigenvalues of the first group (degenerate subspace eigenvalues of S_0).

III. FRACTAL WEYL LAW FOR CNPR

The concept of the fractal Weyl law [31–33] states that the number of states N_λ in a ring of complex eigenvalues with $\lambda_c \leqslant |\lambda| \leqslant 1$ scales in a polynomial way with the growth of matrix size:

$$N_\lambda = a N^b, \tag{5}$$

where the exponent b is related to the fractal dimension of underlying invariant set $d_f = 2b$. The fractal Weyl law was first discussed for the problems of quantum chaotic scattering in the semiclassical limit [31–33]. Later it was shown that this law also works for the Ulam matrix approximant of the Perron-Frobenius operators of dissipative chaotic systems with strange attractors [6,7]. In Ref. [11] it was established that the time growing Linux Kernel network is also characterized by the fractal Weyl law with the fractal dimension $d_f \approx 1.3$.

The fact that $b < 1$ implies that the majority of eigenvalues drop to zero. We see that this property also appears for the CNPR if we test here the validity of the fractal Weyl law by considering a time reduced CNPR of size N_t including the N_t papers published until the time t (measured in years) for

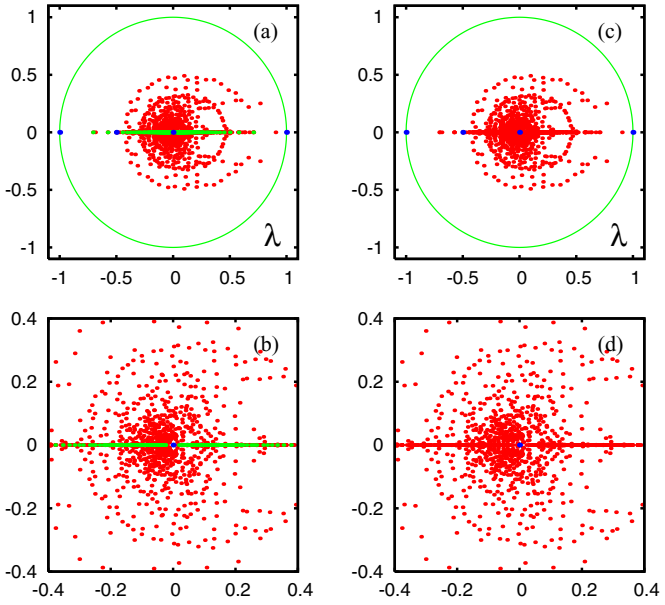

FIG. 10. (Color online) The most accurate spectrum of eigenvalues of S for CNPR. (a) Red/gray dots represent the core space eigenvalues obtained by the rational interpolation method with the numerical precision of $p = 16\,384$ binary digits, $n_R = 2500$ eigenvalues; green (light gray) dots on the $y = 0$ axis show the degenerate subspace eigenvalues of the matrix S_0, which are also eigenvalues of S with a degeneracy reduced by one (eigenvalues of the first group, see text); blue/black dots show the direct subspace eigenvalues of S (same as blue/black dots in left upper panel in Fig. 3). (c) Red/gray dots represent the core space eigenvalues obtained by the high-precision Arnoldi method with $n_A = 2000$ and the numerical precision of $p = 768$ binary digits, and blue dots show the direct subspace eigenvalues of S. Note that the Arnoldi method also determines implicitly the degenerate subspace eigenvalues of S_0, which are therefore not shown in another color. b, d) Same as in top panels (a) and (c) with a zoomed range: $-0.4 \leqslant \mathrm{Re}(\lambda), \mathrm{Im}(\lambda) \leqslant 0.4$.

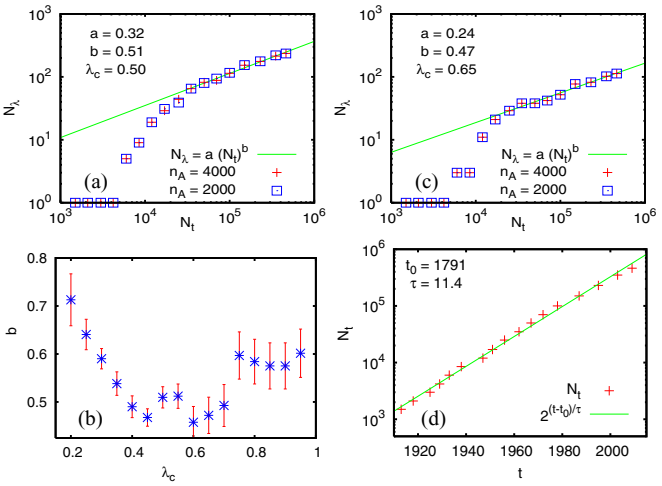

FIG. 11. (Color online) Data for the whole CNPR at different moments of time. (a) [or (c)] Number N_λ of eigenvalues with $\lambda_c \leqslant \lambda \leqslant 1$ for $\lambda_c = 0.50$ (or $\lambda_c = 0.65$) versus the effective network size N_t where the nodes with publication times after a cut time t are removed from the network. The green/gray line shows the fractal Weyl law $N_\lambda = a\,(N_t)^b$ with parameters $a = 0.32 \pm 0.08$ ($a = 0.24 \pm 0.11$) and $b = 0.51 \pm 0.02$ ($b = 0.47 \pm 0.04$) obtained from a fit in the range $3 \times 10^4 \leqslant N_t < 5 \times 10^5$. The number N_λ includes both exactly determined invariant subspace eigenvalues and core space eigenvalues obtained from the Arnoldi method with double precision (52 binary digits) for $n_A = 4000$ (red/gray crosses) and $n_A = 2000$ (blue/black squares). (b) Exponent b with error bars obtained from the fit $N_\lambda = a\,(N_t)^b$ in the range $3 \times 10^4 \leqslant N_t < 5 \times 10^5$ versus cut value λ_c. (d) Effective network size N_t versus cut time t (in years). The green/gray line shows the exponential fit $2^{(t-t_0)/\tau}$ with $t_0 = 1791 \pm 3$ and $\tau = 11.4 \pm 0.2$ representing the number of years after which the size of the network (number of papers published in all Physical Review journals) is effectively doubled.

different times t in order to obtain a scaling behavior of N_λ as a function of N_t. The data presented in Fig. 11 show that the network size grows approximately exponentially as $N_t = 2^{(t-t_0)/\tau}$ with the fit parameters $t_0 = 1791$, $\tau = 11.4$. The time interval considered in Fig. 11 is $1913 \leqslant t \leqslant 2009$ since the first data point corresponds to $t = 1913$ with $N_t = 1500$ papers published between 1893 and 1913. The results for N_λ show that its growth is well described by the relation $N_\lambda = a\,(N_t)^b$ for the range when the number of articles becomes sufficiently large $3 \times 10^4 \leqslant N_t < 5 \times 10^5$. This range is not very large, and probably due to that there is a certain dependence of the exponent b on the range parameter λ_c. At the same time we note that the maximal matrix size N studied here is probably the largest one used in numerical studies of the fractal Weyl law. We have $0.47 < b < 0.6$ for all $\lambda_c \geqslant 0.4$ that is definitely smaller than unity, and thus the fractal Weyl law is well applicable to the CNPR. The value of b increases up to 0.7 for the data points with $\lambda_c < 0.4$, but this is due to the fact the here N_λ also includes some numerically incorrect eigenvalues related to the numerical instability of the Arnoldi method at standard double precision (52 binary digits) as discussed at the beginning of the previous section.

We think that the most appropriate choice for the description of the data is obtained at $\lambda_c = 0.4$, which from one side excludes small, partly numerically incorrect, values of λ and on the other side gives sufficiently large values of N_λ. Here we have $b = 0.49 \pm 02$ corresponding to the fractal dimension $d = 0.98 \pm 0.04$. Furthermore, for $0.4 \leqslant \lambda_c \leqslant 0.7$ we have a rather constant value $b \approx 0.5$ with $d_f \approx 1.0$. Of course, it would be interesting to extend this analysis to a larger size N of CNPR, but for that we still should wait about 10 years until the network size will be doubled compared to the size studied here.

IV. PROPERTIES OF EIGENVECTORS

The results for the eigenvalue spectra of CNPR presented in the previous sections show that most of the visible eigenvalues on the real axis (except for the largest one) in Figs. 9 and 10 are due to the effect of future citations. They appear either directly due to 2×2 subblocks of the type (C2) with a cycle where two papers mutually cite each other giving the degenerate eigenvalues of the first group, or indirectly by eigenvalues of the second group, which are also numerous on the real axis. On the other hand, as can be seen in Fig. 6, for the triangular CNPR, where all future citations are removed, there

is only the leading eigenvalue $\lambda = 1$ and a small number of negative eigenvalues with $-0.27 < \lambda < 0$ on the real axis. All other eigenvalues are complex, and a considerable number of the largest ones are relatively close to corresponding complex eigenvalues for the whole CNPR with future citations.

The appearance of future citations is quite specific and is not a typical situation for citation networks. Therefore we consider the eigenvectors of complex eigenvalues for the triangular CNPR, which indeed represent the typical physical situation without future citations. There is no problem in evaluating these eigenvectors by the Arnoldi method, either with double precision, provided the eigenvalue of the eigenvector is situated in the region of numerically accurate eigenvalues, or with the high-precision variant of the Arnoldi method. However, for the triangular CNPR we have, according to the semianalytical theory presented above, the explicit formula

$$\psi \propto (\lambda \mathbb{1} - S_0)^{-1} e/N = \sum_{j=0}^{l-1} \lambda^{-(1+j)} S_0^j e/N, \qquad (6)$$

where the normalization is given by $\sum_i |\psi(i)| = 1$. This expression is quite convenient, and we verified that it provides the same eigenvectors (up to numerical errors) as the Arnoldi method.

In Fig. 12 we show two eigenvectors of S: one ψ_0 for the leading eigenvalue $\lambda_0 = 1$ and another ψ_{39} for a complex eigenvalue at $|\lambda_{39}| < 1$. The eigenvector of λ_0 gives the PageRank probability for the triangular CNPR (at $\alpha = 1$). We also consider the eigenvector for the complex eigenvalue $\lambda_{39} = -0.3738799 + i\,0.2623941$ (eigenvalues are ordered by their absolute values starting from $\lambda_0 = 1$). In this figure the modulus of $|\psi_j(N_t)|$ is shown versus the time index N_t as introduced in Fig. 11. We also indicate the positions of five famous papers: BCS 1957 [34] at $K = 6$, Anderson 1958 [35] $K = 63$, Benettin *et al.* 1976 [36] $K = 441$, Thouless 1977 [37] $K = 256$, and Abrahams *et al.* 1979 [38] $K = 74$. In the first eigenvector for $\lambda_0 = 1$ all of these papers have quite dominating positions, especially BCS 1957 and Abrahams *et al.* 1979, which are the most important ones if compared to papers of comparable publication date. Only considerably older papers have higher positions in this vector.

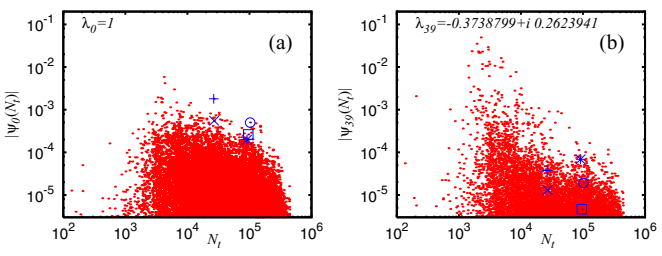

FIG. 12. (Color online) Two eigenvectors of the matrix S for the triangular CNPR. Both panels show the modulus of the eigenvector components $|\psi_j(N_t)|$ versus the time index N_t (as used in Fig. 11) with nodes/articles ordered by the publication time (small red/gray dots). The blue/black points represent five particular articles: BCS 1957 (+), Anderson 1958 (×), Benettin *et al.* 1976 (∗), Thouless 1977 (⊡), and Abrahams *et al.* 1979 (⊙). The left (right) panel corresponds to the real (complex) eigenvalue $\lambda_0 = 1$ ($\lambda_{39} = -0.3738799 + i\,0.2623941$).

For the second eigenvector with complex eigenvalue the older papers (with $10^3 < N_t < 10^4$ corresponding to publications times between 1910 and 1940) are strongly enhanced in its importance, while the above five famous papers lose their importance. The top three positions of largest amplitude $|\psi_{39}(i)|$ correspond to DOI 10.1103/PhysRev.14.409 (1919), 10.1103/PhysRev.8.561 (1916), and 10.1103/PhysRev.24.97 (1917). These old articles study the radiating potentials of nitrogen, ionization impact in gases, and the abnormal low voltage arc, respectively. It is clear that this eigenvector selects a certain community of old articles related to a certain older field of interest. This fact is in agreement with the studies of eigenvectors of Wikipedia network [13] showing that the eigenvectors with $0 < |\lambda| < 1$ select specific communities.

It is interesting to note that the top node of the vector ψ_0 appears in the position $K_{39} = 39$ in local rank index of the vector ψ_{39} (ranking in decreasing order by modulus of $|\psi(i)|$). On the other side the top node of ψ_{39} appears at position $K_0 = 30$ of vector ψ_0. This illustrates how different nodes contribute to different eigenvectors of S.

It is useful to characterize the eigenvectors by their inverse participation ratio (IPR) $\xi_i = (\sum_j |\psi_i(j)|^2)^2 / \sum_j |\psi_i(j)|^4$, which gives an effective number of nodes populated by an eigenvector ψ_i (see, e.g., Refs. [8,13]). For the above two vectors we find $\xi_0 = 20.67$ and $\xi_{39} = 10.76$. This means that ξ_{39} is mainly located on approximately 11 nodes. For ξ_0 this number is twice larger in agreement with data of Fig. 12, which show a clearly broader distribution comparing to ξ_{39}.

We also considered a few tens of eigenstates of S of the whole CNPR. They are mainly located on the complex plane around the largest oval curve well visible in the spectrum (see Fig. 10 top right panel). The IPR value of these eigenstates with $|\lambda| \sim 0.4$ varies in the range $4 < \xi < 13$, showing that they are located on some effective quasi-isolated communities of articles. About 10 of them are related to the top article of ψ_{39} shown in Fig. 12, meaning that these 10 vectors represent various linear combinations of vectors on practically the same community. In global terms, we can say that the eigenstates of G are well localized since $\xi \ll N$. A similar situation was seen for the Wikipedia network [13].

Of course, in addition to ξ it is also useful to consider the whole distribution of ψ amplitudes over the nodes. Such consideration has been done for the Wikipedia network in Ref. [13]. For the CNPR we leave such detailed studies for further investigations.

V. CHEIRANK VERSUS PAGERANK FOR CNPR

The dependence of PageRank probability $P(K)$ on PageRank index K is shown in Fig. 13. The results are similar to those of Refs. [20–23]. We note that the PageRank of the triangular CNPR has the same top nine articles as for the whole CNPR (both at $\alpha = 0.85$ and with a slight interchanged order of positions 7, 8, 9). This confirms that the future citations produce only a small effect on the global ranking.

Following previous studies [25–27], in addition to the Google matrix G we also construct the matrix G^* following the same definition (1) but for the network with inverted direction of links. The PageRank vector of this matrix G^* is called the CheiRank vector with probability $P^*(K_i^*)$ and CheiRank index

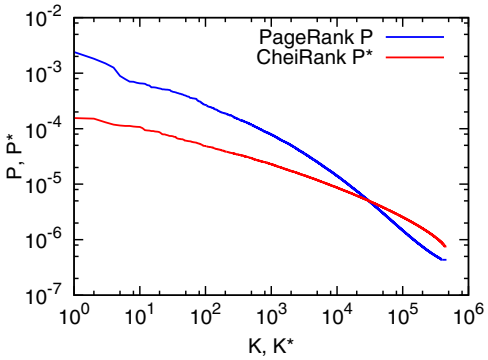

FIG. 13. (Color online) Dependence of probability of PageRank P (CheiRank P^*) on corresponding index K (K^*) for the CNPR at $\alpha = 0.85$.

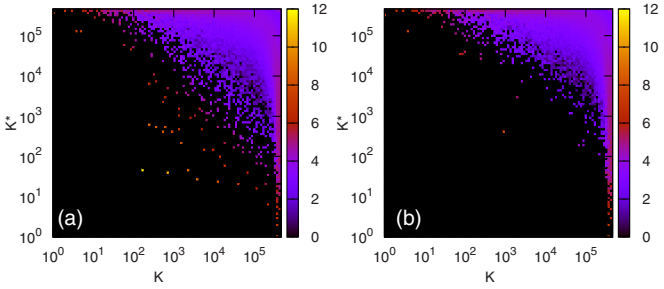

FIG. 14. (Color online) Density distribution $W(K,K^*) = dN_i/dKdK^*$ of Physical Review articles in the PageRank-CheiRank plane (K,K^*). Color bars show the natural logarithm of density, changing from minimal nonzero density (dark) to maximal one (white); zero density is shown by black. (a) All articles of CNPR; (b) CNPR without Rev. Mod. Phys.

K^*. The dependence of $P^*(K_i^*)$ is shown in Fig. 13. We find that the IPR values of P and P^* are $\xi = 59.54$ and 1466.7, respectively. Thus P^* is extended over significantly larger number of nodes comparing to P. A power law fit of the decay $P \propto 1/K^\beta$, $P^* \propto 1/K^{*\beta}$, done for a range $K, K^* \leqslant 2 \times 10^5$, gives $\beta \approx 0.57$ for P and $\beta \approx 0.4$ for P^*. However, this is only an approximate description since there is a visible curvature (in a double logarithmic representation) in these distributions. The corresponding frequency distributions of ingoing links have exponents $\mu = 2.87$, while the distribution of outgoing links has $\mu \approx 3.7$ for out-degree $k \geqslant 20$, even if the whole frequency dependence in this case is rather curved and a power law fit is rather approximate in this case. Thus the usual relation $\beta = 1/(\mu - 1)$ [4,8,26] approximately works.

The correlation between PageRank and CheiRank vectors can be characterized by the correlator $\kappa = N \sum_{i=1}^{N} P(i)P^*(i) - 1$ [25,27]. Here we find $\kappa = -0.2789$ for all CNPR, and $\kappa = -0.3187$ for CNPR without Review of Modern Physics. This is the most strong negative value of κ among all directed networks studied previously [27]. In a certain sense the situation is somewhat similar to the Linux Kernel network where $\kappa \approx 0$ or slightly negative ($\kappa > -0.1$ [25]). For CNPR, we can say that due to an almost triangular structure of G and G^* there is a very little overlap of top ranking in K and K^* that leads to a negative correlator value, since the components $P(i)P^*(i)$ of the sum for κ are small.

Each article i has two indexes K_i, K_i^* so that it is convenient to see their distribution on a 2D PageRank-CheiRank plane. The density distribution $W(K,K^*) = dN_i/dKdK^*$ is shown in Fig. 14. It is obtained from 100×100 cells equidistant in log scale (see details in Refs. [26,27]). For the CNPR the density is homogeneous along lines $K = -K^* + \text{const}$, which corresponds to the absence of correlations between P and P^* [26,27]. For the CNPR without Review of Modern Physics we have an additional suppression of density at low K^* values. Indeed, Review of Modern Physics contains mainly review articles with a large number of citations that place them on top of CheiRank. At the top three positions of K^* of CNPR we have DOI 10.1103/PhysRevA.79.062512, 10.1103/PhysRevA.79.062511, and 10.1103/RevModPhys.81.1551 of 2009. These are articles with long citation lists on K shell diagram $4d$ transition elements, hypersatellites of $3d$ transition metals, and superconducting phases of f electron compounds, respectively. For CNPR without Review of Modern Physics the first two articles are the same, and the third one has DOI 10.1103/PhysRevB.80.224501 being about models for the coexistence of d wave superconducting and charge density wave order in high-temperature cuprate superconductors. We see that the most recent articles with long citation lists are dominating.

The top PageRank articles are analyzed in detail in Refs. [20–23], and we do not discuss them here.

It is also useful to consider two-dimensional rank 2DRank K_2 defined by counting nodes in order of their appearance on ribs of squares in (K,K^*) plane with the square size growing from $K = 1$ to N [26]. It selects highly cited articles with a relatively long citation list. For CNPR, we have the top three such articles with DOI 10.1103/RevModPhys.54.437 (1982), 10.1103/RevModPhys.65.851 (1993), and 10.1103/RevModPhys.58.801 (1986). Their topics are electronic properties of two-dimensional systems, pattern formation outside of equilibrium, and spin glasses facts and concepts, respectively. The first one located at $K = 183$, $K^* = 49$ is well visible in the left panel of Fig. 14. For CNPR without Review of Modern Physucs we find at $K_2 = 1$ the article with DOI 10.1103/PhysRevD.54.1 (1996) entitled "Review of particle physics" with much information on physical constants.

For the ranking of articles about persons in Wikipedia networks [14,26,39], PageRank, 2DRank, and CheiRank highlight in a different manner various aspects of human activity. For the CNPR, these three ranks also select different types of articles; however, due to a triangular structure of G, G^* and absence of correlations between PageRank and CheiRank vectors, the useful side of 2DRank and CheiRank remains less evident.

VI. IMPACTRANK FOR INFLUENCE PROPAGATION

It is interesting to quantify how an influence of a given article propagates through the whole CNPR. To analyze this property we consider the following propagator acting on an initial vector v_0 located on a given article:

$$v_f = \frac{1-\gamma}{1-\gamma G} v_0, \quad v_f^* = \frac{1-\gamma}{1-\gamma G^*} v_0. \tag{7}$$

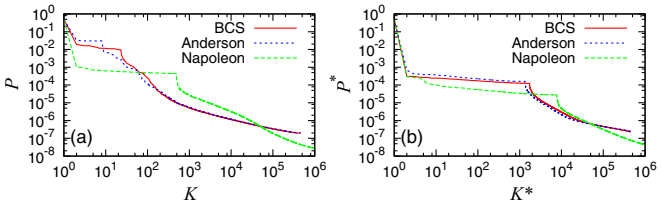

FIG. 15. (Color online) Dependence of impact vector v_f probability P and P^* (a, b) on the corresponding ImpactRank index K and K^* for an initial article v_0 as BCS [34] and Anderson [35] in CNPR, and Napoleon in the English Wikipedia network from Ref. [39]. Here the impact damping factor is $\gamma = 0.5$.

Here G, G^* are the Google matrices defined above, γ is a new impact damping factor in a range $\gamma \sim 0.5$–0.9, and v_f in the final vector generated by the propagator (7). This vector is normalized to unity $\sum_i v_f(i) = 1$, and one can easily show that it is equal to the PageRank vector of a modified Google matrix given by

$$\tilde{G} = \gamma\, G + (1-\gamma)\, v_0\, e^T, \tag{8}$$

where e is the vector with unit elements. This modified Google matrix corresponds to a stochastic process where at a certain time a given probability distribution is propagated with probability γ using the initial Google matrix G, and with probability $(1-\gamma)$ the probability distribution is reinitialized with the vector v_0. Then v_f is the stationary vector from this stochastic process. Since the initial Google matrix G has a similar form, $G = \alpha S + (1-\alpha) e\, e^T/N$ with the damping factor α, the modified Google matrix can also be written as

$$\tilde{G} = \tilde{\alpha}\, S + (1-\tilde{\alpha})\, v_p\, e^T, \quad \tilde{\alpha} = \gamma\alpha, \tag{9}$$

with the personalization vector [4]

$$v_p = \frac{\gamma(1-\alpha)e/N + (1-\gamma)v_0}{1-\gamma\alpha}, \tag{10}$$

which is also sum normalized: $\sum_i v_p(i) = 1$. Obviously similar relations hold for G^* and v_f^*.

The relation (7) can be viewed as a Green function with damping γ. Since $\gamma < 1$ the expansion in a geometric series is convergent, and v_f can be obtained from about 200 terms of the expansion for $\gamma \sim 0.5$. The stability of v_f is verified by changing the number of terms. The obtained vectors v_f, v_f^* can be considered as effective PageRank, CheiRank probabilities P, P^*, and all nodes can be ordered in the corresponding rank index K, K^*, which we will call ImpactRank.

The results for two initial vectors located on BCS [34] and Anderson [35] articles are shown in Fig. 15. In addition we show the same probability for the Wikipedia article "Napoleon" for the the English Wikipedia network analyzed in Ref. [39]. The direct analysis of the distributions shows that the original article is located at the top position, and the next steplike structure corresponds to the articles reached by first outgoing (ingoing) links from v_0 for G (G^*). The next visible step correspond to a second link step.

The top 10 articles for these three vectors are shown in Tables I, II, III, IV, and V. The analysis of these top articles confirms that they are closely linked with the initial article, and thus the ImpactRank gives relatively good ranking results. At the same time, some questions for such ImpactRanking still remain to be clarified. For example, in Table IV we find at the third position the well-known Review of Modern Physics article on Anderson transitions, but the paper of Abrahams *et al.* [38] appears only on far positions $K^* \approx 300$. The situation is changed if we consider all CNPR links as bidirectional, obtaining a nondirectional network. Then the paper [38] appears on the second position directly after initial article [35]. We think that such a problem appears due to triangular structure of CNPR, where there is no intersection of forward and backward flows. Indeed, for the case of Napoleon we do not see such difficulties. Thus we hope that such an approach can be applied to other directed networks.

VII. MODELS OF RANDOM PERRON-FROBENIUS MATRICES

In this section we discuss the spectral properties of several random matrix models of Perron-Frobenius operators characterized by non-negative matrix elements and column sums normalized to unity. We call these models random Perron-Frobenius matrices (RPFM). To construct these models for a given matrix G of dimension N we draw N^2 independent matrix elements $G_{ij} \geqslant 0$ from a given distribution $p(G)$ [with $p(G) = 0$ for $G < 0$] with average $\langle G \rangle = 1/N$ and finite variance $\sigma^2 = \langle G^2 \rangle - \langle G \rangle^2$. A matrix obtained in this way obeys the column sum normalization only in average but not exactly for an arbitrary realization. Therefore

TABLE I. Spreading of impact of "Theory of superconductivity" paper by J. Bardeen, L. N. Cooper, and J. R. Schrieffer (doi:10.1103/PhysRev.108.1175) by Google matrix G with $\alpha = 0.85$ and $\gamma = 0.5$.

ImpactRank	DOI	Title of paper
1	10.1103/PhysRev.108.1175	Theory of superconductivity
2	10.1103/PhysRev.78.477	Isotope effect in the superconductivity of mercury
3	10.1103/PhysRev.100.1215	Superconductivity at millimeter wave frequencies
4	10.1103/PhysRev.78.487	Superconductivity of isotopes of mercury
5	10.1103/PhysRev.79.845	Theory of the superconducting state. I. The ground ...
6	10.1103/PhysRev.80.567	Wave functions for superconducting electrons
7	10.1103/PhysRev.79.167	The hyperfine structure of Ni^{61}
8	10.1103/PhysRev.97.1724	Theory of the Meissner effect in superconductors
9	10.1103/PhysRev.81.829	Relation between lattice vibration and London ...
10	10.1103/PhysRev.104.844	Transmission of superconducting films ...

TABLE II. Spreading of impact of "Absence of diffusion in certain random lattices" paper by P. W. Anderson (doi:10.1103/PhysRev.109.1492) by Google matrix G with $\alpha = 0.85$ and $\gamma = 0.5$.

ImpactRank	DOI	Title of paper
1	10.1103/PhysRev.109.1492	Absence of diffusion in certain random lattices
2	10.1103/PhysRev.91.1071	Electronic structure of f centers: Saturation of ...
3	10.1103/RevModPhys.15.1	Stochastic problems in physics and astronomy
4	10.1103/PhysRev.108.590	Quantum theory of electrical transport phenomena
5	10.1103/PhysRev.48.755	Theory of pressure effects of foreign gases on spectral lines
6	10.1103/PhysRev.105.1388	Multiple scattering by quantum-mechanical systems
7	10.1103/PhysRev.104.584	Spectral diffusion in magnetic resonance
8	10.1103/PhysRev.74.206	A note on perturbation theory
9	10.1103/PhysRev.70.460	Nuclear induction
10	10.1103/PhysRev.90.238	Dipolar broadening of magnetic resonance lines ...

we renormalize all columns to unity after having drawn the matrix elements. This renormalization provides some (hopefully small) correlations between the different matrix elements.

Neglecting these correlations for sufficiently large N the statistical average of the RPFM is simply given by $\langle G_{ij}\rangle = 1/N$, which is a projector matrix with the eigenvalue $\lambda = 1$ of multiplicity 1 and the corresponding eigenvector being the uniform vector e (with $e_i = 1$ for all i). The other eigenvalue $\lambda = 0$ is highly degenerate of multiplicity $N - 1$, and its eigenspace contains all vectors orthogonal to the uniform vector e. Writing the matrix elements of a RPFM as $G_{ij} = \langle G_{ij}\rangle + \delta G_{ij}$ we may consider the fluctuating part δG_{ij} as a perturbation, which only weakly modifies the unperturbed eigenvector e for $\lambda = 1$, but for the eigenvalue $\lambda = 0$ we have to apply degenerate perturbation theory which requires the diagonalization of δG_{ij}. According to the theory of nonsymmetric real random Gaussian matrices [5,40,41] it is well established that the complex eigenvalue density of such a matrix is uniform on a circle of radius $R = \sqrt{N}\sigma$ with σ^2 being the variance of the matrix elements. One can also expect that this holds for more general, non-Gaussian, distributions with finite variance provided that we exclude extreme long tail distribution where the typical values are much smaller than σ. Therefore we expect that the eigenvalue density of a RPFM is determined by a single parameter being the variance σ^2 of the matrix elements, resulting in a uniform density on a circle of radius $R = \sqrt{N}\sigma$ around $\lambda = 0$, in addition to the unit eigenvalue $\lambda = 1$, which is always an exact eigenvalue due to sum normalization of columns.

We now consider different variants of RPFM. The first variant is a full matrix with each element uniformly distributed in the interval $[0,2/N[$, which gives the variance $\sigma^2 = 1/(3N^2)$ and the spectral radius $R = 1/\sqrt{3N}$. The second variant is a sparse RPFM matrix with Q nonvanishing elements per column and which are uniformly distributed in the interval $[0,2/Q[$. Then the probability distribution is given by $p(G) = (1 - Q/N)\delta(G) + (Q/N)\,\chi_{[0,2/Q[}(G)$ where $\chi_{[0,2/Q[}(G)$ is the characteristic function on the interval $[0,2/Q[$ (with values being 1 for G in this interval and 0 for G outside this interval). The average is indeed $\langle G\rangle = 1/N$, and the variance is $\sigma^2 = 4/(3NQ)$ (for $N \gg Q$) providing the spectral radius $R = 2/\sqrt{3Q}$. We may also consider a sparse RPFM where we have exactly Q nonvanishing constant elements of value $1/Q$ in each column with random positions resulting in a variance $\sigma^2 = 1/(NQ)$ and $R = 1/\sqrt{Q}$. The theoretical predictions for these three variants of RPFM coincide very well with numerical simulations. In Fig. 16 the complex eigenvalue spectrum for one realization of each of the three cases is shown for $N = 400$ and $Q = 20$, clearly confirming the circular uniform eigenvalue density with the theoretical values of R. We also confirm numerically the scaling behavior of R as a function of N or Q.

Motivated by the Google matrices of DNA sequences [42], where the matrix elements are distributed with a power law, we also considered a power law variant of RPFM with

TABLE III. Spreading of impact of "Theory of superconductivity" paper by J. Bardeen, L. N. Cooper, and J. R. Schrieffer (doi:10.1103/PhysRev.108.1175) by Google matrix G^* with $\alpha = 0.85$ and $\gamma = 0.5$.

ImpactRank	DOI	Title of paper
1	10.1103/PhysRev.108.1175	Theory of superconductivity
2	10.1103/PhysRevB.77.104510	Temperature-dependent gap edge in strong-coupling ...
3	10.1103/PhysRevC.79.054328	Exact and approximate ensemble treatments of thermal ...
4	10.1103/PhysRevB.8.4175	Ultrasonic attenuation in superconducting molybdenum
5	10.1103/RevModPhys.62.1027	Properties of boson-exchange superconductors
6	10.1103/PhysRev.188.737	Transmission of far-infrared radiation through thin films ...
7	10.1103/PhysRev.167.361	Superconducting thin film in a magnetic field—Theory of ...
8	10.1103/PhysRevB.77.064503	Exact mesoscopic correlation functions of the Richardson ...
9	10.1103/PhysRevB.10.1916	Magnetic field attenuation by thin superconducting lead films
10	10.1103/PhysRevB.79.180501	Exactly solvable pairing model for superconductors with ...

TABLE IV. Spreading of impact of "Absence of diffusion in certain random lattices" paper by P. W. Anderson (doi:10.1103/PhysRev.109.1492) by Google matrix G^* with $\alpha = 0.85$ and $\gamma = 0.5$.

ImpactRank	DOI	Title of paper
1	10.1103/PhysRev.109.1492	Absence of diffusion in certain random lattices
2	10.1103/PhysRevA.80.053606	Effects of interaction on the diffusion of atomic ...
3	10.1103/RevModPhys.80.1355	Anderson transitions
4	10.1103/PhysRevE.79.041105	Localization-delocalization transition in hessian ...
5	10.1103/PhysRevB.79.205120	Statistics of the two-point transmission at ...
6	10.1103/PhysRevB.80.174205	Localization-delocalization transitions ...
7	10.1103/PhysRevB.80.024203	Statistics of renormalized on-site energies and ...
8	10.1103/PhysRevB.79.153104	Flat-band localization in the Anderson-Falicov-Kimball model
9	10.1103/PhysRevB.74.104201	One-dimensional disordered wires with Poschl-Teller potentials
10	10.1103/PhysRevB.71.235112	Critical wave-packet dynamics in the power-law bond ...

$p(G) = D(1 + aG)^{-b}$ for $0 \leqslant G \leqslant 1$ and with an exponent $2 < b < 3$. The condition $G \leqslant 1$ is required because of the column sum normalization. The parameters D and a are determined by normalization and the average $\langle G \rangle = 1/N$. In the limit $N^{b-2} \gg 1$ we find $a \approx N/(b-2)$ and $D \approx N(b-1)/(b-2)$. For $b > 3$ the variance would scale with $\sim N^{-2}$ resulting in $R \sim 1/\sqrt{N}$ as in the first variant with uniformly distributed matrix elements. However, for $b < 3$ this scaling is different, and we find (for $N^{b-2} \gg 1$)

$$R = C(b)\, N^{1-b/2}, \quad C(b) = (b-2)^{(b-1)/2} \sqrt{\frac{b-1}{3-b}}. \qquad (11)$$

Figure 17 shows the results of numerical diagonalization for one realization with $N = 400$ and $b = 2.5$ such that we expect $R \sim N^{-0.25}$. It turns out that the circular eigenvalue density is rather well confirmed and the "theoretical radius" is indeed given by $R = \sqrt{N}\sigma$ if the variance σ^2 of matrix elements is determined by an average over the N^2 matrix elements of the given matrix. A study for different values of N with $50 \leqslant N \leqslant 2000$ also confirms the dependence $R = C\, N^{-\eta}$ with fit values $C = 0.67 \pm 0.03$ and $\eta = 0.22 \pm 0.01$. The value of $\eta = 0.22$ is close to the theoretical value $1 - b/2 = 0.25$ but the prefactor $C = 0.67$ is smaller than its theoretical value $C(2.5) \approx 1.030$. This is due to the correlations introduced by the additional column sum normalization after drawing the random matrix elements. Furthermore for the power law model with $b < 3$ we should not expect a precise confirmation of the uniform circular density obtained for Gaussian distribution matrix elements. Actually, a more detailed numerical analysis of the density shows that the density for the power law model is not exactly uniform, in particular for values of b close to 2.

The important observation is that a generic RPFM (full, sparse, or with power law distributed matrix elements) has a complex eigenvalue density rather close to a uniform circle of a quite small radius (depending on the parameters N, Q, or b). The fact, that the realistic networks (e.g., certain university WWW networks) have Google matrix spectra very different from this [10] shows that in these networks there is indeed a subtle network structure and that slight random perturbations or variations already immediately result in uniform circular eigenvalue spectra. This was already observed in Refs. [8,9], where it was shown that certain modest random changes in the network links already provide such circular eigenvalue spectra.

We also determine the PageRank for the different variants of the RPFM, i.e., the eigenvector for the eigenvalue $\lambda = 1$. It turns out that PageRank vector has practically equal probabilities on each node. This is natural since this eigenvector should be close to the uniform vector e, which is the "PageRank" for the average matrix $\langle G_{ij} \rangle = 1/N$. This also holds when we use a damping factor $\alpha = 0.85$ for the RPFM.

Following the above discussion about triangular networks (with $G_{ij} = 0$ for $i \geqslant j$) we also study numerically a triangular RPFM where for $j \geqslant 2$ and $i < j$ the matrix elements G_{ij} are uniformly distributed in the interval $[0, 2/(j-1)[$, and for $i \geqslant j$ we have $G_{ij} = 0$. Then the first column is empty, which means it corresponds to a dangling node and it needs to be replaced by $1/N$ entries. For the triangular RPFM the situation

TABLE V. Spreading of impact of the article "Napoleon" in English Wikipedia by Google matrix G and G^* with $\alpha = 0.85$ and $\gamma = 0.5$.

ImpactRank	Articles (G case)	Articles (G^* case)
1	Napoleon	Napoleon
2	French Revolution	List of orders of battle
3	France	Lists of state leaders by year
4	First French Empire	Names inscribed under the Arc de Triomphe
5	Napoleonic Wars	List of battles involving France
6	French First Republic	Order of battle of the Waterloo Campaign
7	Saint Helena	Napoleonic Wars
8	French Consulate	Wagram order of battle
9	French Directory	Departments of France
10	National Convention	Jena-Auerstedt Campaign Order of Battle

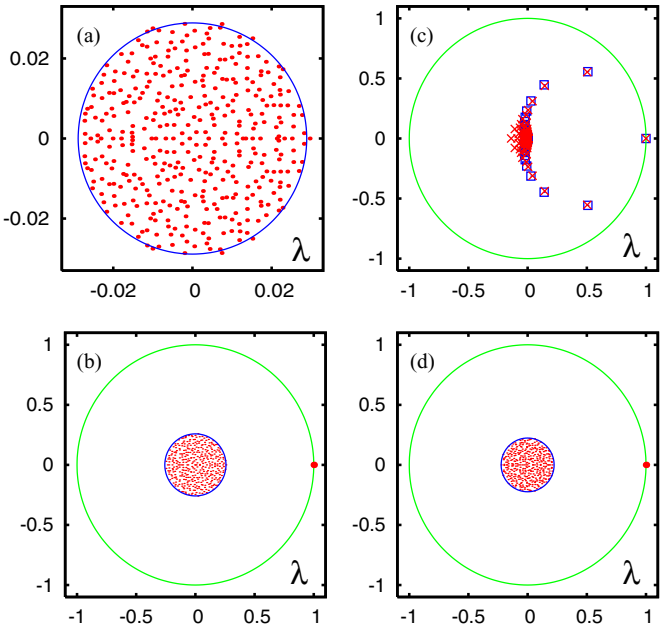

FIG. 16. (Color online) (a) Spectrum (red/gray dots) of one realization of a full uniform RPFM with dimension $N = 400$ and matrix elements uniformly distributed in the interval $[0,2/N[$; the blue/black circle represents the theoretical spectral border with radius $R = 1/\sqrt{3N} \approx 0.02887$. The unit eigenvalue $\lambda = 1$ is not shown due to the zoomed presentation range. (c) Spectrum of one realization of triangular RPFM (red/gray crosses) with nonvanishing matrix elements uniformly distributed in the interval $[0,2/(j-1)[$ and a triangular matrix with nonvanishing elements $1/(j-1)$ (blue/black squares); here $j = 2,3,\ldots,N$ is the index number of nonempty columns and the first column with $j = 1$ corresponds to a dangling node with elements $1/N$ for both triangular cases. (b, d) Complex eigenvalue spectrum (red/gray dots) of a sparse RPFM with dimension $N = 400$ and $Q = 20$ nonvanishing elements per column at random positions. Panel (b) [or (d)] corresponds to the case of uniformly distributed nonvanishing elements in the interval $[0,2/Q[$ (constant nonvanishing elements being $1/Q$); the blue/black circle represents the theoretical spectral border with radius $R = 2/\sqrt{3Q} \approx 0.2582$ ($R = 1/\sqrt{Q} \approx 0.2236$). In panels (b) and (d) $\lambda = 1$ is shown by a larger red dot for better visibility. The unit circle is shown by green/gray line [panels (b), (c), and (d)].

changes completely since here the average matrix $\langle G_{ij}\rangle = 1/(j-1)$ (for $i < j$ and $j \geqslant 2$) has already a nontrivial structure and eigenvalue spectrum. Therefore the argument of degenerate perturbation theory which allowed us to apply the results of standard full nonsymmetric random matrices does not apply here. In Fig. 16 one clearly sees that for $N = 400$ the spectra for one realization of a triangular RPFM and its average are very similar for the eigenvalues with large modulus but both do not have at all a uniform circular density in contrast to the RPRM models without the triangular constraint discussed above. For the triangular RPFM the PageRank behaves as $P(K) \sim 1/K$ with the ranking index K being close to the natural order of nodes $\{1,2,3,\ldots\}$, which reflects the fact that the node 1 has the maximum of $N-1$ incoming links and so on.

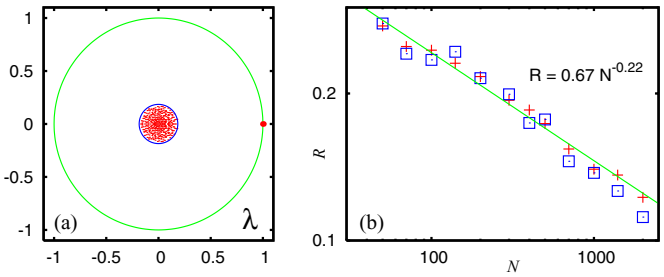

FIG. 17. (Color online) (a) Spectrum (red/gray dots) of one realization of the power law RPFM with dimension $N = 400$ and decay exponent $b = 2.5$ (see text); the unit eigenvalue $\lambda = 1$ is shown by a large red/gray dot, the unit circle is shown by a green/gray curve; the blue/black circle represents the spectral border with theoretical radius $R = \approx 0.1850$ (see text). (b) Dependence of the spectrum border radius on matrix size N for $50 \leqslant N \leqslant 2000$; red/gray crosses represent the radius obtained from theory (see text); blue/black squares correspond to the spectrum border radius obtained numerically from a small number of eigenvalues with maximal modulus; the green/gray line shows the fit $R = C\,N^{-\eta}$ of red/gray crosses with $C = 0.67 \pm 0.03$ and $\eta = 0.22 \pm 0.01$.

The study of above models shows that it is not so simple to find a good RPFM model which reproduces a typical spectral structure of real directed networks.

VIII. DISCUSSION

In this study we presented a detailed analysis of the spectrum of the CNPR for the period 1893–2009. It happens that the numerical simulations should be done with a high accuracy (up to $p = 16\,384$ binary digits for the rational interpolation method or $p = 768$ binary digits for the high-precision Arnoldi method) to determine correctly the eigenvalues of the Google matrix of CNPR at small eigenvalues λ. Due to the time ordering of citations, the CNPR G matrix is close to the triangular form with a nearly nilpotent matrix structure. We show that special semianalytical methods allow us to determine efficiently the spectrum of such matrices. The eigenstates with large modulus of λ are shown to select specific communities of articles in certain research fields, but there is no clear way on how to identify a community one is interested in.

The obtained results show that the spectrum of CNPR is characterized by the fractal Weyl law with the fractal dimension $d_f \approx 1$ and the growth exponent $b \approx 0.5$ being significantly smaller than unity. We think that the Physical Review network has a structure which is typical for other citation networks, and thus our result shows that the fractal Weyl law is a typical feature of citation networks.

The ranking of articles is analyzed with the help of PageRank and CheiRank vectors corresponding to forward and backward citation flows in time. It is shown that the correlations between these two vectors are small and even negative, which is similar to the case of Linux Kernel networks [27] and significantly different from networks of universities and Wikipedia. The 2DRanking on the PagRank-CheiRank plane allows us to select articles which efficiently redistribute information flow on the CNPR.

To characterize the local impact propagation for a given article we introduce the concept of ImpactRank, which efficiently determines its domain of influence.

Finally we perform the analysis of several models of RPFM showing that such full random matrices are very far from the realistic cases of directed networks. Random sparse matrices with a limited number Q of links per nodes seem to be closer to typical Google matrices concerning the matrix structure. However, such random models give a rather uniform eigenvalue density with a spectral radius $\sim 1/\sqrt{Q}$ and a flat PageRank distribution. Furthermore they do not capture the existence of quasi-isolated communities, which generates a quasidegenerate spectrum at $\lambda = 1$. Further development of RPFM models is required to reproduce the spectral properties of real modern directed networks.

In summary, we developed powerful numerical methods which allowed us to determine numerically the exact eigenvalues and eigenvectors of the Google matrix of Physical Review. We demonstrated that this matrix is close to triangular matrices of large size where numerical errors can significantly affect the eigenvalues. We show that the techniques developed in this work allow us to resolve such difficulties and obtain the exact spectrum in a semianalytical manner. The eigenvectors of eigenvalues with $|\lambda| < 1$ are located on certain communities of articles related to specific scientific research subjects. We point that the random matrix models of Google matrices are still awaiting their detailed development. Indeed, matrices with random elements have a spectrum being very different from the real one. Thus, while the random matrix theory of Hermitian and unitary matrices has been very successful (see, e.g., Ref. [5]), a random matrix theory for Google matrices still waits its development. It is possible that the case of triangular matrices, which is rather similar to our CNPR case, can be a good starting point for development of such models.

ACKNOWLEDGMENTS

We thank the American Physical Society for letting us use their citation database for Physical Review [16]. This research is supported in part by the EC FET Open project "New tools and algorithms for directed network analysis" (NADINE No. 288956). This work was granted access to the HPC resources of CALMIP (Toulouse) under the allocation 2012-P0110.

APPENDIX A: THEORY OF TRIANGULAR ADJACENCY MATRICES

Let us briefly mention the analytical theory of Ref. [19] for pure triangular networks with a nilpotent matrix S_0 such that $S_0^l = 0$. For integers [19] the adjacency matrix is defined as $A_{mn} = k$ where k is a multiplicity defined as the largest integer such that m^k is a divisor of n and if $1 < m < n$, and $k = 0$ if $m = 1$ or $m = n$ or if m is not a divisor of n. Thus, we have $k = 0$ if m is not a divisor of n and $k \geqslant 1$ if m is a divisor of n different from 1 and n. The total size N of the matrix is fixed by the maximal considered integer. Then the Google matrix is constructed from A_{mn} following the standard rules described above. This network of integers gives an important example of a triangular Google matrix with similar features also appearing in the Physical Review citation network. Below we discuss the general properties of such matrices.

For this we define the coefficients:

$$c_j = d^T S_0^j\, e/N, \quad b_j = e^T S_0^j\, e/N, \tag{A1}$$

which are nonzero only for $j = 0, 1, \ldots, l-1$. The fact that the nonvanishing columns of S_0 are sum normalized and that the other columns (corresponding to dangling nodes) are zero can be written as $e^T S_0 = e^T - d^T$ implying $d^T = e^T(\mathbb{1} - S_0)$. Using this identity and the fact that $S_0^k = 0$ for $k \geqslant l$ we find

$$\sum_{k=j}^{l-1} c_k = d^T(\mathbb{1} - S_0)^{-1} S_0^j\, e/N = e^T S_0^j\, e/N = b_j, \tag{A2}$$

and in particular for $j = 0$ we obtain the sum rule $\sum_{k=0}^{l-1} c_k = 1$ and for $j = l-1$ the identity $b_{l-1} = c_{l-1}$.

Consider now a right eigenvector ψ of S with eigenvalue λ. If $d^T\psi = 0$ we find from (4) that ψ is also an eigenvector of S_0, and since S_0 is nilpotent the eigenvalue must be $\lambda = 0$. Therefore for $\lambda \neq 0$ we have necessarily $d^T\psi \neq 0$, and with the appropriate normalization of ψ we have $d^T\psi = 1$, which implies together with the eigenvalue equation $\psi = (\lambda\mathbb{1} - S_0)^{-1}\, e/N$ where the matrix inverse is well defined for $\lambda \neq 0$. The eigenvalue is determined by the condition

$$0 = \lambda^l(1 - d^T\psi) = \lambda^l\left(1 - d^T \frac{\mathbb{1}}{\lambda\mathbb{1} - S_0} e/N\right). \tag{A3}$$

Since S_0 is nilpotent we may expand the matrix inverse in a finite series, and therefore the eigenvalue λ is the zero of the reduced polynomial of degree l:

$$\mathcal{P}_r(\lambda) = \lambda^l - \sum_{j=0}^{l-1} \lambda^{l-1-j}\, c_j, \tag{A4}$$

where the coefficients c_j are given by (A1). Using $d^T = e^T(\mathbb{1} - S_0)$ we may rewrite (A3) in the form

$$0 = \lambda^l\left(1 - e^T \frac{\mathbb{1} - S_0}{\lambda\mathbb{1} - S_0} e/N\right) = (\lambda - 1)\lambda^l\, e^T \frac{\mathbb{1}}{\lambda\mathbb{1} - S_0} e/N, \tag{A5}$$

which gives another expression for the reduced polynomial:

$$\mathcal{P}_r(\lambda) = (\lambda - 1)\sum_{j=0}^{l-1} \lambda^{l-1-j}\, b_j, \tag{A6}$$

using the coefficients b_j and confirming explicitly that $\lambda = 1$ is indeed an eigenvalue of S. The expression (A6) can also be obtained by a direct calculation from (A2) and (A4).

Since the reduced polynomial has at most l zeros λ_j ($\neq 0$ since $c_{l-1} = b_{l-1} \neq 0$) we find that there are at most l nonvanishing eigenvalues of S given by these zeros. They can also be obtained as the eigenvalues of a "small" $l \times l$ matrix. To see this let us define the following set of vectors v_j for $j = 1, \ldots, l$ by $v_j = c_{j-1}^{-1} S_0^{j-1}\, e/N$, where we have chosen to apply the prefactor c_{j-1}^{-1} to the vector $S_0^{j-1}\, e/N$ [43]. From (4) and (A1) one finds that Sv_j can be expanded in the other

vectors v_k as

$$S v_j = \frac{c_j}{c_{j-1}} v_{j+1} + c_0 v_1 = \sum_{k=1}^{l} \bar{S}_{kj} v_k, \qquad \text{(A7)}$$

where $\bar{S}_{kj}$ are the matrix elements of the $l \times l$ representation matrix

$$\bar{S} = \begin{pmatrix} c_0 & c_0 & \cdots & c_0 & c_0 \\ c_1/c_0 & 0 & \cdots & 0 & 0 \\ 0 & c_2/c_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & c_{l-1}/c_{l-2} & 0 \end{pmatrix}. \qquad \text{(A8)}$$

Note that for the last vector v_l we have $S v_l = c_0 v_1$ since $c_l = 0$, and therefore the matrix $\bar{S}$ provides a closed and mathematically exact representation of S on the l-dimensional subspace generated by $v_1, \ldots, v_l$. Furthermore one can easily verify (by a recursive calculation in l) that the characteristic polynomial of $\bar{S}$ coincides with the reduced polynomial (A4). Therefore numerical diagonalization of $\bar{S}$ provides an alternative method to compute the nonvanishing eigenvalues of S. In principle one can also determine directly the zeros of the reduced polynomial by the Newton-Maehly method, and in Ref. [19] this was indeed done for cases with very modest values of $l \leqslant 29$. However, here for the triangular CNPR we have $l = 352$ and the coefficients c_j become very small, especially $c_{l-1} \approx 3.6 \times 10^{-352}$, a number which is (due to the exponent) outside the range of 64-bit standard double-precision numbers (IEEE 754) with 52 bits for the mantissa, 10 bits for the exponent (with respect to 2), and two bits for the signs of mantissa and exponent. This exponent range problem is not really serious and can, for example, be circumvented by a smart reformulation of the algorithm to evaluate the ratio $\mathcal{P}_r(\lambda)/\mathcal{P}'_r(\lambda)$ using only ratios c_j/c_{j-1}, which do not have this exponent range problem. However, it turns out that in this approach the convergence of the Newton-Maehly method using double-precision arithmetic is very bad for many zeros and does not provide reliable results. In Appendix B we show how this problem can be solved using high-precision calculations, but we mention that one may also try another approach by diagonalizing numerically the representation matrix $\bar{S}$ given in (A8), which also depends on the ratios c_j/c_{j-1}.

APPENDIX B: EFFECTS OF NUMERICAL ERRORS AND HIGH-PRECISION COMPUTATIONS

We note that the Arnoldi method determines an orthonormal set of vectors $\zeta_1, \zeta_2, \zeta_3, \ldots, \zeta_{n_A}$ where the first vector ζ_1 is obtained by normalizing a given initial vector and ζ_{j+1} is obtained by orthonormalizing $S\zeta_j$ to the previous vectors determined so far. It is obvious due to (A7) that for the initial uniform vector e each ζ_j is given by a linear combination of the vectors v_k with $k = 1, \ldots, j$. Since the subspace of v_k for $k = 1, \ldots, l$ is closed with respect to applications of S the Arnoldi method should, in theory, break off at $n_A = l$ with a zero coupling element. The latter is given as the norm of $S\zeta_l$ othogonalized to $\zeta_1, \ldots, \zeta_l$, and if this norm vanishes the vector ζ_{l+1} cannot be constructed and the Arnoldi method has completely explored an S-invariant subspace of dimension l.

However, due to a strong effect of round-off errors and the fact that the vectors v_j are numerically "nearly" linearly dependent the last coupling element does not vanish numerically (when using double precision) and the Arnoldi method produces a cloud of numerically incorrect eigenvalues due to the Jordan blocks, which are mathematically outside the representation space (defined by the vectors v_j) but which are still explored due to round-off errors and clearly visible in Fig. 5. The double-precision spectrum of $\bar{S}$ seems to provide well-defined eigenvalues in the range where the Arnoldi method produces the "Jordan block cloud," but outside this cloud both spectra coincide only partly, mainly for the eigenvalues with a largest modulus and positive real part. For the eigenvalues with a negative real part there are considerable deviations. As can be seen in Fig. 6 the eigenvalues produced by the Arnoldi method at double precision are reliable provided that they are well *outside* the Jordan block cloud of incorrect eigenvalues. Therefore the deviations outside the Jordan block cloud show that the numerical double-precision diagonalization of the representation matrix $\bar{S}$ is not reliable as well, but here the effect of numerical errors is quite different as for the Arnoldi method, as is explained below.

In order to obtain an alternative and reliable numerical method to determine the spectrum of the triagonal CNRP we have also tried to determine the zeros of the reduced polynomial using higher-precision numbers with 80 or even 128 bits (quadruple precision), which helps to solve the (minor) exponent range problem (mentioned in Appendix A) because these formats use more bits for the exponent. However, there are indeed two other serious numerical problems. First, it turns out that in a certain range of the complex plane around $\mathrm{Re}(\lambda) \approx -0.1$ to -0.2 and $\mathrm{Im}(\lambda) \leqslant 0.1$ the numerical evaluation of the polynomial suffers in a severe way from an alternate sign problem with a strong loss of significance. Second, the zeros of the polynomial depend in a very sensitive way on the precision of the coefficients c_j (see below). We have found that even 128-bit numbers are not sufficient to obtain all zeros with a reasonable graphical precision.

Therefore we use the very efficient GNU Multiple Precision Arithmetic Library (GMP library) [30]. With this library one has 31 bits for the exponent, and one may chose an arbitrary number of bits for the mantissa. We find that using 256 bits (binary digits) for the mantissa the complex zeros of the reduced polynomial can be determined with a precision of 10^{-18}. In this case the convergence of the Newton-Maehly method is very nice, and we find that the sum (and product) of the complex zeros coincide with a high precision with the theoretical values c_0 [respectively: $(-1)^{l-1} c_{l-1}$] due to (A4). We have also tested different ways to evaluate the polynomial, such as the Horner scheme versus direct evaluation of the sum and for both methods using both expressions (A4) and (A6). It turns out that with 256 binary digits during the calculation the zeros obtained by the different variants of the method coincide very well within the required precision of 10^{-18}. Of course, the coefficients c_j or b_j given by (A1) also need to be evaluated with the precision of 256 binary digits, but there is no problem of using high-precision vectors since the nonvanishing matrix elements of S_0 are rational numbers,

which allow us to perform the evaluation of the vectors $S_0^j e/N$ with arbitrary precision. We also tested a random modification of c_j according to $c_j \to c_j(1 + 10^{-16}X)$ where X is a random number in the interval $]-0.5, 0.5[$. This modification gives significant differences of the order of 10^{-2} to 10^{-1} for some of the complex zeros and which are very visible in the graphical representation of the spectra. Therefore, the spectrum depends in a very sensitive way on these coefficients, and it is now quite clear that numerical double-precision diagonalization of $\bar{S}$, which depends according to (A8) on the values c_j, cannot provide accurate eigenvalues simply because the double-precision round-off errors of c_j imply a sensitive change of eigenvalues. In particular some of the numerical eigenvalues of $\bar{S}$ differ quite strongly from the high-precision zeros of the reduced polynomial.

In order to study more precisely the effect of the numerical instability of the Arnoldi method due to the Jordan blocks we also use the GMP library to increase the numerical precision of the Arnoldi method. To be precise we implement the first part of this method, the *Arnold iteration*, in which the $n_A \times n_A$ Arnoldi representation matrix is determined by the Gram-Schmidt orthogonalization procedure, using high-precision numbers, while for the second step, the numerical diagonalization of this representation matrix, we keep the standard double precision. It turns that only the first step is numerically critical. Once the Arnoldi representation matrix is obtained in a careful and precise way, it is numerically well conditioned, and its numerical diagonalization works well with only double precision.

APPENDIX C: THEORY OF DEGENERATE EIGENVALUES

In order to understand the mechanism of the degenerate core space eigenvalues visible in Fig. 8 we extend the argumentation of Appendix A for triangular CNPR to the case of nearly triangular networks. Consider again the matrix S given by Eq. (4), but now S_0 is not nilpotent. There are two groups of eigenvectors ψ of S with eigenvalue λ. The first group is characterized by the orthogonality $d^T\psi = 0$ of the eigenvector ψ with respect to the dangling vector d, and the second group is characterized by the nonorthogonality $d^T\psi \neq 0$. In the following, we describe efficient methods to determine all eigenvalues of the first group and a considerable number of eigenvalues of the second group. We note that for the case of a purely triangular network the first group contains only eigenvectors for the eigenvalue 0, and the second group contains the eigenvectors for the l nonvanishing eigenvalues as discussed in Appendix A. In principle there are also complications due to generalized eigenvectors (associated to nontrivial Jordan blocks), but they appear mainly for the eigenvalue zero, and for the moment we do not discuss these complications.

First, we note that the subspace eigenvectors of S belong to the first group because the nodes of the subspaces of S cannot contain dangling nodes, which, by construction of S, are linked to any other node and therefore belong to the core space. Since any subspace eigenvector ψ has nonvanishing values only for subspace nodes being different from dangling nodes we have obviously $d^T\psi = 0$. We also note that an eigenvector of S of the first group with $d^T\psi = 0$ is due to (4) also an eigenvector of S_0 with the same eigenvalue.

For the remaining eigenvectors in the first group one might try to diagonalize the matrix S_0 and check for each eigenvector of S_0 if the identity $d^T\psi = 0$ holds, in which case we would obtain an eigenvector of S of the first group but generically, and apart from the subspace eigenvectors, there is no reason that eigenvectors of S_0 with isolated nondegenerate eigenvalues obey this identity. However, if we have an eigenvalue of S_0 with a degeneracy $m \geqslant 2$ we may construct by suitable linear combinations $m-1$ linearly independent eigenvectors of S_0 which also obey $d^T\psi = 0$, and therefore this eigenvalue with degeneracy m of S_0 is also an eigenvalue with degeneracy $m-1$ of S. In order to determine the degenerate eigenvalues of S_0 it is useful to determine the subspaces of S_0, which (in contrast to the subspaces of S) may contain dangling nodes. Actually, each dangling node is a trivial invariant subspace (for S_0) of dimension 1 with a network matrix of size 1×1 and being zero. Explicitly we have implemented the following procedure: first, we determine the subspaces of S (with 71 nodes in total) and remove these nodes from the network. Then we determine all subspaces of S_0 whose dimension is below 10. Each time such a subspace is found its nodes are immediately removed from the network. When we have tested in a first run all nodes as potential subspace nodes the procedure is repeated until no new subspaces of maximal dimension 10 are found since removal of former subspaces may have created new subspaces. Then the limit size of 10 is doubled to 20, 40, 80, etc., to ensure that we do not miss large subspaces. However, for the CNPR it turns out that the limit size of 10 allows us to find all subspaces. In our procedure a subsequently found subspace may potentially have links to a former subspace leading to a block-triangular structure (and not block-diagonal structure as was done in Ref. [10]). This method to determine "relative" subspaces of a network already reduced by former subspaces is more convenient for the CNPR, which is nearly triangular, and it allows us also to determine correctly all subspace eigenvalues by diagonalizing each relative subspace network. The removal of subspace nodes of S and S_0 reduces the network size from $N = 463\,348$ to $404\,959$. In the next step we remove in the same way the subspaces of the transpose S_0^T of S_0 (since the eigenvalues of S_0^T and S_0 are the same), which reduces the network size further to $90\,965$. In total this procedure provides a block-triangular structure of S_0 as

$$S_0 = \begin{pmatrix} S_1 & * & \cdots & & & \cdots & * \\ 0 & S_2 & * & & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & & \vdots \\ & & 0 & B & * & & \\ \vdots & & & 0 & T_1 & * & \vdots \\ \vdots & & & & 0 & T_2 & * \\ 0 & \cdots & & & \cdots & \ddots & \ddots \end{pmatrix}, \quad \text{(C1)}$$

where $S_1, S_2, \ldots$ represent the diagonal subblocks associated to the subspaces of S and S_0 and $T_1, T_2, \ldots$ represent the

diagonal subblocks associated to the subspaces of S_0^T, and B is the "bulk" part for the remaining network of 90 965 nodes. The stars represent potential nonvanishing entries whose values do not influence the eigenvalues of S_0. The subspace blocks $S_1, S_2, \ldots$ and $T_1, T_2, \ldots$, which are individually of maximal dimension 10, can be directly diagonalized, and it turns that out of 372 382 eigenvalues in these blocks only about 4000 eigenvalues (counting degeneracies) or 950 eigenvalues (noncounting degeneracies) are different from zero. Most of these eigenvalues are not degenerate and are therefore not eigenvalues of S, but there are still quite many degenerate eigenvalues at $\lambda = \pm 1/\sqrt{n}$ with $n \geqslant 2$ taking small integer values and that are also eigenvalues of S with a degeneracy reduced by one.

Concerning the bulk block B we can write it in the form $B = B_0 + f_1 e_1^T$, where f_1 is the first column vector of B and $e_1^T = (1, 0, \ldots, 0)$. The matrix B_0 is obtained from B by replacing its first column to zero. We can apply the above argumentation between S and S_0 in the same way to B and B_0; i.e., the degenerate eigenvalues of B_0 with degeneracy m are also eigenvalues of B with degeneracy $m-1$ (with eigenvectors obeying $e_1^T \psi = 0$) and therefore eigenvalues of S with degeneracy $m-2$. The matrix B_0 is decomposed in a similar way as in (C1) with subspace blocks, which can be diagonalized numerically, and a new bulk block $\tilde{B}$ of dimension 63 559 and which may be treated in the same way by taking out its first column. This procedure provides a recursive scheme which after nine iterations stops with a final bulk block of zero size. At each iteration we keep only subspace eigenvalues with degeneracies $m \geqslant 2$ and which are joined with reduced degeneracies $m-1$ to the subspace spectrum of the previous iteration. For this joined spectrum we keep again only eigenvalues with degeneracies $m \geqslant 2$ which are joined with the subspace spectrum of the next higher level and so on.

In this way we have determined all eigenvalues of S_0 with a degeneracy $m \geqslant 2$ which belong to the eigenvalues of S of the first group. Including the direct subspace of S there are 4999 nonvanishing eigenvalues (counting degeneracies) or 442 nonvanishing eigenvalues (noncounting degeneracies). The degeneracy of the zero eigenvalue (or the dimension of the generalized kernel) is found by this procedure to be 455 789, but this would only be correct assuming that there are no general eigenvectors of higher order (representation vectors of nontrivial Jordan blocks), which is clearly not the case. The Jordan subspace structure of the zero eigenvalue complicates the argumentation. Here at each iteration step the degeneracy has to be reduced from m to $m-D$ where $D > 1$ is the dimension of the maximal Jordan block since each generalized eigenvector at a given order has to be treated as an independent vector when constructing vectors obeying the orthogonality with respect to the dangling vector d. Therefore the degeneracy of the zero eigenvalue cannot be determined exactly, but we may estimate its degeneracy of about $\sim 455\,000$ out of 463 348 nodes in total. This implies that the number of nonvanishing eigenvalues is about $\sim$8000–9000, which is considerably larger than the value of 352 for the triangular CNPR but still much smaller than the total network size.

In Table VI we provide the degeneracies for some of the eigenvalues $\pm 1/\sqrt{n}$ for integer n in the range $1 \leqslant n \leqslant 25$.

TABLE VI. Degeneracies of the eigenvalues with largest modulus for the whole CNPR whose eigenvectors ψ belong to the first group and obey the orthogonality $d^T \psi = 0$ with the dangling vector d.

λ	Degeneracy
1	27
-1	18
$\pm 1/\sqrt{2}$	27
$\pm 1/\sqrt{3}$	20
1/2	58
$-1/2$	52
$\pm 1/\sqrt{5}$	20
$\pm 1/\sqrt{6}$	52
$\pm 1/\sqrt{7}$	6
$\pm 1/\sqrt{8}$	44
1/3	47
$-1/3$	39
$\pm 1/\sqrt{10}$	33
$\pm 1/\sqrt{11}$	1
$\pm 1/\sqrt{12}$	85
$\pm 1/\sqrt{14}$	15
$\pm 1/\sqrt{15}$	46
1/4	52
$-1/4$	42
$\pm 1/\sqrt{18}$	29
$\pm 1/\sqrt{20}$	60
$\pm 1/\sqrt{21}$	30
$\pm 1/\sqrt{22}$	3
$\pm 1/\sqrt{24}$	69
1/5	20
$-1/5$	11

The degeneracies for $+1/\sqrt{n}$ and $-1/\sqrt{n}$ are identical for nonsquare numbers n (with noninteger $\sqrt{n}$) and different for square numbers (with integer $\sqrt{n}$). Apparently for nonsquare numbers the eigenvalues are only generated from effective 2×2 blocks:

$$\begin{pmatrix} 0 & 1/n_1 \\ 1/n_2 & 0 \end{pmatrix} \quad \Rightarrow \lambda = \pm \frac{1}{\sqrt{n_1 n_2}} \tag{C2}$$

with positive integers n_1 and n_2 such that $n = n_1 n_2$, while for square numbers $n = m^2$ they may be generated by such blocks or by simple 1×1 blocks containing $1/m$ such that the degeneracy for $+1/\sqrt{n} = +1/m$ is larger than the degeneracy for $-1/\sqrt{n} = -1/m$. Furthermore, statistically the degeneracy is smaller for prime numbers n or numbers with less factorization possibilities and larger for numbers with more factorization possibilities. The Arnoldi method (with 52 binary digits for double-precision arithmetic and $n_A = 8000$) provides according to the sizes of the plateaux visible in Fig. 8 the overall approximate degeneracies $\sim$60 for $|\lambda| = 1/\sqrt{2}$ (i.e., $\pm 1/\sqrt{2}$ counted together), $\sim$50 for $|\lambda| = 1/\sqrt{3}$, and $\sim$ 115 for $|\lambda| = 1/2$. These values are coherent with (but slightly larger than) the values 54, 40, and 110 taken from Table VI. Actually, as we will see below, the slight differences between the degeneracies obtained from Fig. 8 and from Table VI are

indeed relevant and correspond to some eigenvalues of the second group which are close but not identical to $\pm 1/\sqrt{2}$, $\pm 1/\sqrt{3}$, or $\pm 1/2$ and do not contribute in Table VI.

APPENDIX D: RATIONAL INTERPOLATION METHOD

We now consider the eigenvalues λ of S for the eigenvectors of the second group with nonorthogonality $d^T\psi \neq 1$ or $d^T\psi = 1$ after proper renormalization of ψ. Now ψ cannot be an eigenvector of S_0, and λ is not an eigenvalue of S_0. Similarly as in Appendix A the eigenvalue equation $S\psi = \lambda\psi$, the condition $d^T\psi = 1$ and (4) imply that the eigenvalue λ of S is a zero of the rational function

$$\mathcal{R}(\lambda) = 1 - d^T \frac{\mathbb{1}}{\lambda\mathbb{1} - S_0} e/N = 1 - \sum_{j,q} \frac{C_{jq}}{(\lambda - \rho_j)^q}, \qquad \text{(D1)}$$

where we have formally expanded the vector e/N in eigenvectors of S_0 and with ρ_j being the eigenvalues of S_0 and q is the *order* of the eigenvector of ρ_j used in this expansion, i.e., $q = 1$ for simple eigenvectors and $q > 1$ for generalized eigenvectors of higher order due to Jordan blocks. Note that even the largest possible value of q for a given eigenvalue may be (much) smaller than its multiplicity m. Furthermore the case of simple repeating eigenvalues (with simple eigenvectors) with higher multiplicity $m > 1$ leads only to several identical terms $\sim (\lambda - \rho_j)^{-1}$ for any eigenvector of this eigenvalue, thus all contributing to the coefficients C_{jq} and whose precise values we do not need to know in the following. For us the important point is that the second identity in (D1) establishes that $\mathcal{R}(\lambda)$ is indeed a rational function whose denominator and numerator polynomials have the same degree and whose poles are (some of) the eigenvalues of S_0.

We mention that one can also show by a simple determinant calculation (similar to a calculation shown in Ref. [19] for triangular networks with nilpotent S_0) that

$$P_S(\lambda) = P_{S_0}(\lambda)\,\mathcal{R}(\lambda), \qquad \text{(D2)}$$

where $P_S(\lambda)$ [or $P_{S_0}(\lambda)$] is the characteristic polynomial of S (S_0). Therefore those zeros of $\mathcal{R}(\lambda)$ which are not zeros of $P_{S_0}(\lambda)$ (i.e., not eigenvalues of S_0) are indeed zeros of $P_S(\lambda)$ (i.e., eigenvalues of S) since there are not poles of $\mathcal{R}(\lambda)$. Furthermore, generically the *simple* zeros $P_{S_0}(\lambda)$ also appear as poles in $\mathcal{R}(\lambda)$ and are therefore not eigenvalues of S. However, for a zero of $P_{S_0}(\lambda)$ (eigenvalue of S_0) with *higher multiplicity* $m > 1$ (and unless m is equal to the maximal Jordan block order q associated to this eigenvalue of S_0) the corresponding pole in $\mathcal{R}(\lambda)$ only reduces the multiplicity to $m - 1$ (or $m - q$ in case of higher order generalized eigenvectors), and we also have a zero of $P_S(\lambda)$ (eigenvalue of S). Some of the eigenvalues of S_0, whose eigenvectors ψ are orthogonal to the dangling vector ($d^T\psi = 0$) and do not contribute in the expansion in (D1), are not poles of $\mathcal{R}(\lambda)$ and therefore also eigenvalues of S. This concerns essentially the direct subspace eigenvalues of S, which are also direct subspace eigenvalues of S_0 as already discussed in Appendix C. In total the identity (D2) confirms exactly the above picture that there are two groups of eigenvalues and with the special role of direct subspace eigenvalues belonging to the first group.

Our aim is to determine numerically the zeros of the rational function $\mathcal{R}(\lambda)$. In order to evaluate this function we expand the first identity in (D1) in a matrix geometric series, and we obtain

$$\mathcal{R}(\lambda) = 1 - \sum_{j=0}^{\infty} c_j \lambda^{-1-j} \qquad \text{(D3)}$$

with the coefficients c_j defined in (A1) and provided that this series converges. In Appendix A, where we discussed the case of a nilpotent matrix S_0 with $S_0^l = 0$, the series was finite, and for this particular case we had $\mathcal{R}(\lambda) = \lambda^{-l}\mathcal{P}_r(\lambda)$ where $\mathcal{P}_r(\lambda)$ was the reduced polynomial defined in (A4) and whose zeros provided the l nonvanishing eigenvalues of S for nilpotent S_0.

However, for the CNPR the series are infinite since all c_j are different from zero. One may first try a crude approximation and simply replace the series by a finite sum for $j < l$ and using some rather large cutoff value for l and determine the zeros in the same way as for the nilpotent case (a high-precision calculation of the zeros of the reduced polynomial of degree l). It turns that in this way we obtain correctly the largest core space eigenvalue of S as $\lambda_1 = 0.999\,751\,822\,283\,878$ which is also obtained by (any variant of) the Arnoldi method. However, the other zeros obtained by this approximation lie all on a circle of radius ≈ 0.9 in the complex plane and obviously do not represent any valid eigenvalues. Increasing the cutoff value l does not help either, and it increases only the density of zeros on this circle. To understand this behavior we note that in the limit $j \to \infty$ the coefficients c_j behave as $c_j \propto \rho_1^j$ where $\rho_1 = 0.902\,448\,280\,519\,224$ is the largest eigenvalue of the matrix S_0 with an eigenvector nonorthogonal to d. Note that the matrix S_0 also has some degenerate eigenvalues at $+1$ and -1, but these eigenvalues are obtained from the direct subspace eigenvectors of S (which are also direct subspace eigenvectors of S_0) and which are orthogonal to the dangling vector d and do not contribute in the rational function (D1). It turns actually out that the eigenvalue ρ_1 is also the largest *subspace space* eigenvalue of S_0 (after having removed the direct subspace nodes of S). By analyzing explicitly the small-dimensional subspace related to this eigenvalue one can show that ρ_1 is given as the largest solution of the polynomial equation $x^3 - \frac{2}{3}x - \frac{2}{15} = 0$ and can therefore be expressed as $\rho_1 = 2\,\mathrm{Re}\,[(9 + i\sqrt{119})^{1/3}]/(135)^{1/3}$. The asymptotic behavior $c_j \propto \rho_1^j$ is also confirmed by the direct numerical evaluation of c_j. Therefore the series (D3) converges only for $|\lambda| > \rho_1$, and a simple (even very large) cutoff in the sum implies that only eigenvalues $|\lambda_j| > \rho_1$ can be determined as a zero of the finite sum. The only eigenvalue respecting this condition is the largest core space eigenvalue λ_1 given above.

One may try to improve this by a "better" approximation, which consists of evaluating the sum exactly up to some value l and then to replace the remaining sum as a geometric series with the approximation $c_j \approx c_l \rho_1^{j-l}$ for $j \geqslant l$ and with ρ_1 determined as the ratio $\rho_1 = c_l/c_{l-1}$ (which provides a sufficient approximation) or taken as its exact (high precision) value. This improved approximation results in $\mathcal{R}(\lambda) \approx \lambda^{-l}(\lambda - \rho_1)^{-1}\mathcal{P}(\lambda)$ with a polynomial $\mathcal{P}(\lambda)$ whose zeros provide in total four correct eigenvalues. Apart from λ_1 it also gives $\lambda_2 = 0.902\,445\,536\,212\,661$ (note that this eigenvalue of S is very close but different to the eigenvalue ρ_1 of S_0) and $\lambda_{3,4} = 0.765\,857\,950\,563\,684 \pm i\,0.251\,337\,495\,625\,571$ such that $|\lambda_{3,4}| = 0.806\,045\,245\,100\,386$. All these four core space

eigenvalues coincide very well with the first four eigenvalues obtained from the Arnoldi method. However, the other zeros of the polynomial $\mathcal{P}(\lambda)$ lie again on a circle, now with a reduced radius ≈ 0.7, and do not coincide with eigenvalues of S. This can be understood by the fact that the coefficients c_j obey for $j \to \infty$ the more precise asymptotic expression $c_j \approx C_1\rho_1^j + C_2\rho_2^j + C_2\rho_3^j + \cdots$ with the next eigenvalues $\rho_2 = 1/\sqrt{2} \approx 0.707$ and $\rho_3 = -\rho_2$. Here the first term $C_1\rho_1^j$ is dealt with analytically by the replacement of the geometric series, but the other terms create a new convergence problem. Therefore the improved approximation allows us only to determine the four core space eigenvalues with $|\lambda_j| > |\rho_{2,3}| = 1/\sqrt{2}$. To obtain more valid eigenvalues it seems to be necessary to sum up by geometric series many of the next terms, not only the next two terms due to ρ_2 and ρ_3, but also the following terms of smaller eigenvalues ρ_j of S_0. In other words the exact pole structure of the rational function $\mathcal{R}(\lambda)$ has be kept as best as possible.

Therefore due to the rational structure of the function $\mathcal{R}(\lambda)$ with many eigenvalues ρ_j of S_0 that determine its precise pole structure we suggest the following numerical approach using high-precision arithmetic. For a given number p of binary digits, e.g., $p = 1024$, we determine the coefficients c_j for $j < l$ where the cutoff value

$$l \approx \frac{\ln(1-\rho_1) - p\ln(2)}{\ln(\rho_1)} \approx 6.753\, p + \text{const} \qquad \text{(D4)}$$

is sufficiently large to evaluate the sum (D3) accurately in the given precision of p binary digits (error below 2^{-p}) for *all complex values* λ *on the unit circle*, i.e., $|\lambda| = 1$, where the series converges well. Furthermore we choose a number n_R of "eigenvalues" we want to calculate, e.g., $n_R = 300$, and evaluate the rational function $\mathcal{R}(z)$ at $n_S = 2n_R + 1$ support points $z_j = \exp(2\pi i\, j/n_S)$ $(j = 0, \ldots, n_S - 1)$ uniformly distributed on the unit circle using the series (D3). Then we calculate the rational function $R_I(z)$ which interpolates $\mathcal{R}(z)$ at the n_S support points z_j, $R_I(z_j) = \mathcal{R}(z_j)$, using Thiele's interpolation formula. Then the numerator and denominator polynomials of $R_I(z)$ are both of degree n_R. Thiele's interpolation formula expresses $R_I(z)$ in terms of a continued fraction expansion using inverse differences. This method is quite standard and well described in the literature of numerical mathematics; see, for example, Ref. [44]. After having evaluated a table of n_S inverse differences (with $n_S^2/2$ operations) one can evaluate arbitrary values of $R_I(z)$ using the continued fraction expansion (with n_S operations). It is not very difficult to derive from the continued fraction expansion a recursive scheme to evaluate the values of the numerator and denominator polynomials separately as well as their derivatives. Using this scheme we determine the n_R complex zeros of the numerator polynomial using the (high-precision variant of the) Newton-Maehly method. These zeros correspond to the zeros of the rational functional $\mathcal{R}(z)$ and are taken as approximate eigenvalues of the matrix S of the second group. The main idea of this approach is to evaluate these zeros from the analytical continuation of $\mathcal{R}(z)$ using values for $|z| = 1$ to determine its zeros well inside the unit circle.

We also consider a second variant of the method where the number of support points $n_S = 2n_R + 2$ is even (instead of $n_S = 2n_R + 1$ being odd as for the first variant). In this case the numerator polynomial is of degree $n_R + 1$ (instead of n_R) while the denominator polynomial is of degree n_R, and we choose to interpolate the inverse of the rational function $1/\mathcal{R}(z)$ [instead of $\mathcal{R}(z)$ itself] by $R_I(z)$ such that the zeros of $\mathcal{R}(z)$ are given by the n_R zeros of the denominator (instead of the numerator) polynomial of $R_I(z)$.

The number n_R must not be too small in order to well approximate the second identity in (D1) by the fit function. On the other hand for a given precision of p binary digits the number of n_R must not be too large as well because the coefficients c_j, which may be written as the expansion $c_j = \sum_\nu C_\nu\, \rho_\nu^j$, do not contain enough information to resolve its structure for the smaller eigenvalues ρ_j of S_0. Therefore for too large values of n_R (for a given precision), we obtain additional artificial zeros of the numerator polynomial (or of the denominator polynomial for the second variant) of $R_I(z)$, mostly close to the unit circle, somehow as additional nodes around the support points.

It turns out that for the proper combination of p and n_R values the method provides highly accurate eigenvalues and works astonishingly well. In particular for values of n_R below a certain threshold (depending on the precision p) both variants of the method with odd or even number of support points provide numerically identical zeros (with final results rounded to 52 binary digits), which indeed coincide very accurately (for most of them) with the eigenvalues of S we want to determine.

We note that the rational interpolation method allows us only to determine the eigenvalues of S of the second group, i.e., the eigenvalues which are not eigenvalues of S_0 and whose eigenvectors obey $d^T\psi \neq 0$. The eigenvalues of the first group (with $d^T\psi = 0$) have to be determined separately by the scheme of degenerate subspace eigenvalues of S_0 described in Appendix C. In particular the eigenvalues given in Table VI and belonging to the first group are not zeros of the rational function $\mathcal{R}(z)$ (they are actually poles of this function), but it turns out that there are some zeros of $\mathcal{R}(z)$ which are very close but not identical to some of the values in Table VI. For example, the rational interpolation method provides the following zeros: $1/2 + 3.134\,010\,98 \times 10^{-5}$, $1/2 + 1.327\,9300 \times 10^{-7}$, $1/\sqrt{2} - 1.1597 \times 10^{-10}$ or $1/\sqrt{2} - 6.419\,004 \times 10^{-8}$, which are indeed accurate in the given precision since they are stable for all values of $p \geqslant 1024$ and the corresponding maximal value of n_R, and we have stopped the Newton iteration when the error of a zero was clearly below 10^{-18}. These zeros are also found with the same precision in the data of the high-precision Arnoldi method for the three different values of 256, 512, or 768 binary digits. However, based only on results of the Arnoldi method it is not really clear if the small corrections to $1/2$ or $1/\sqrt{2}$ are real and exact or numerically artificial since the Arnoldi method has indeed problems with degenerate and clustered eigenvalues [17]. Therefore the rational interpolation method provides an independent and strong confirmation of the accuracy of these type of eigenvalues. We attribute their existence to a quasisubspace structure, similarly as discussed in Ref. [10], with a matrix subblock as in (C2) but which is still very weakly coupled (by many indirect network links) to the core space.

[1] S. Brin and L. Page, Comput. Netw. ISDN Syst. **30**, 107 (1998).

[2] A. A. Markov, *Rasprostranenie zakona bol'shih chisel na velichiny, zavisyaschie drug ot druga*, Izvestiya Fiziko-matematicheskogo obschestva pri Kazanskom universitete, 2-ya seriya, **15**, 135 (1906) (in Russian) [English trans.: *Extension of the limit theorems of probability theory to a sum of variables connected in a chain*, reprinted in Appendix B of R. A. Howard, *Dynamic Probabilistic Systems*, Vol. 1, *Markov Models* (Dover Publications, New York, 2007)].

[3] M. Brin and G. Stuck, *Introduction to Dynamical Systems* (Cambridge University Press, Cambridge, 2002).

[4] A. M. Langville and C. D. Meyer, *Google's PageRank and Beyond: The Science of Search Engine Rankings* (Princeton University Press, Princeton, 2006).

[5] M. L. Mehta, *Random Matrices* (Elsevier-Academic Press, Amsterdam, 2004).

[6] D. L. Shepelyansky and O. V. Zhirov, Phys. Rev. E **81**, 036213 (2010).

[7] L. Ermann and D. L. Shepelyansky, Eur. Phys. J. B **75**, 299 (2010).

[8] O. Giraud, B. Georgeot, and D. L. Shepelyansky, Phys. Rev. E **80**, 026107 (2009).

[9] B. Georgeot, O. Giraud, and D. L. Shepelyansky, Phys. Rev. E **81**, 056109 (2010).

[10] K. M. Frahm, B. Georgeot, and D. L. Shepelyansky, J. Phys, A: Math. Theor. **44**, 465101 (2011).

[11] L. Ermann, A. D. Chepelianskii, and D. L. Shepelyansky, Eur. Phys. J. B **79**, 115 (2011).

[12] K. M. Frahm and D. L. Shepelyansky, Eur. Phys. J. B **85**, 355 (2012).

[13] L. Ermann, K. M. Frahm, and D. L. Shepelyansky, Eur. Phys. J. B **86**, 193 (2013).

[14] Y.-H. Eom, K. M. Frahm, A. Benczur, and D. L. Shepelyansky, Eur. Phys. J. B **86**, 492 (2013).

[15] R. Albert and A.-L. Barabási, Phys. Rev. Lett. **85**, 5234 (2000).

[16] Web page of Physical Review: http://publish.aps.org/.

[17] G. W. Stewart, *Matrix Algorithms, Eigensystems* (SIAM, Washington, DC, 2001), Vol. II.

[18] K. M. Frahm and D. L. Shepelyansky, Eur. Phys. J. B **76**, 57 (2010).

[19] K. M. Frahm, A. D. Chepelianskii, and D. L. Shepelyansky, J. Phys. A: Math. Theor. **45**, 405101 (2012).

[20] S. Redner, Phys. Today **58**(6), 49 (2005).

[21] P. Chen, H. Xie, S. Maslov, and S. Redner, J. Infometrics **1**, 8 (2007).

[22] F. Radicchi, S. Fortunato, B. Markines, and A. Vespignani, Phys. Rev. E **80**, 056103 (2009).

[23] Y.-H. Eom and S. Fortunato, PLoS ONE **6**, e24926 (2011).

[24] J. D. West, T. C. Bergstrom, and C. T. Bergstrom, Coll. Res. Libr. **71**, 236 (2010); http://www.eigenfactor.org/.

[25] A. D. Chepelianskii, arXiv:1003.5455 [cs.Se] (2010).

[26] A. O. Zhirov, O. V. Zhirov, and D. L. Shepelyansky, Eur. Phys. J. B **77**, 523 (2010).

[27] L. Ermann, A. D. Chepelianskii, and D. L. Shepelyansky, J. Phys. A: Math. Theor. **45**, 275101 (2012).

[28] This number depends on the exact time ordering which is used and which is not unique because many papers are published at the same time and the order between them is not specified. We have chosen a time ordering where between these papers, degenerate in publication time, the initial node order of the raw data is kept.

[29] Note that some of the nonvanishing components of the iteration vector $S_0^i e$ may become very small, e.g., $\sim 10^{-100}$. In this context we count such components still as occupied despite their small size, and N_i is the number of nodes which can be reached from some arbitrary other node after i iterations with the matrix S_0.

[30] T. Granlund and the GMP development team, http://gmplib.org/.

[31] J. Sjöstrand, Duke Math. J. **60**, 1 (1990).

[32] J. Sjöstrand and M. Zworski, Duke Math. J. **137**, 381 (2007).

[33] S. Nonnenmacher and M. Zworski, Commun. Math. Phys. **269**, 311 (2007).

[34] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. **108**, 1175 (1957).

[35] P. W. Anderson, Phys. Rev. **109**, 1492 (1958).

[36] G. Benettin, L. Galgani, and J.-M. Strelcyn, Phys. Rev. A **14**, 2338 (1976).

[37] D. J. Thouless, Phys. Rev. Lett. **39**, 1167 (1977).

[38] E. Abrahams, P. W. Anderson, D. C. Licciardello, and T. V. Ramakrishnan, Phys. Rev. Lett. **42**, 673 (1979).

[39] Y.-H. Eom and D. L. Sepelyansky, PLoS ONE **8**, e74554 (2013).

[40] J. Ginibre, J. Math. Phys. Sci. **6**, 440 (1965).

[41] H.-J. Sommers, A. Crisanti, H. Sompolinsky, and Y. Stein, Phys. Rev. Lett. **60**, 1895 (1988); N. Lehmann and H.-J. Sommers, *ibid.* **67**, 941 (1991).

[42] V. Kandiah and D. L. Shepelyansky, PLoS ONE **8**, e61519 (2013).

[43] In Ref. [19] a set of vectors without this prefactor was used, but this provided a representation matrix which is numerically unstable for a direct diagonalization. The prefactor c_{j-1}^{-1} ensures that the representation matrix is numerically (rather) stable, and of course both matrices are mathematically related by a similarity transformation and have identical eigenvalues.

[44] J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis* (Springer, New York, 2002).

Google matrix analysis of directed networks

Leonardo Ermann

Departamento de Física Teórica, GIyA, Comisión Nacional de Energía Atómica, Buenos Aires, Argentina

Klaus M. Frahm and Dima L. Shepelyansky

Laboratoire de Physique Théorique du CNRS, IRSAMC, Université de Toulouse, UPS, 31062 Toulouse, France

(Dated: July 25, 2014)

In past ten years, modern societies developed enormous communication and social networks. Their classification and information retrieval processing become a formidable task for the society. Due to the rapid growth of World Wide Web, social and communication networks, new mathematical methods have been invented to characterize the properties of these networks on a more detailed and precise level. Various search engines are essentially using such methods. It is highly important to develop new tools to classify and rank enormous amount of network information in a way adapted to internal network structures and characteristics. This review describes the Google matrix analysis of directed complex networks demonstrating its efficiency on various examples including World Wide Web, Wikipedia, software architecture, world trade, social and citation networks, brain neural networks, DNA sequences and Ulam networks. The analytical and numerical matrix methods used in this analysis originate from the fields of Markov chains, quantum chaos and Random Matrix theory.

Keywords: Markov chains, World Wide Web, search engines, complex networks, PageRank, 2DRank, CheiRank

"The Library exists *ab aeterno*."
Jorge Luis Borges *The Library of Babel*

Contents

I. INTRODUCTION

On a scale of ten years, modern societies developed enormous communication and social networks. The World Wide Web (WWW) alone has about 50 billion indexed web pages, so that their classification and information retrieval processing become a formidable task which the society has to face every day. Various search engines have been developed by private companies such as Google, Yahoo! and others which are extensively used by Internet users. In addition, social networks (Facebook, LiveJournal, Twitter, etc) gained enormous popularity in the last few years. Active use of social networks spreads beyond their initial purposes making them important for political or social events.

To handle such enormous databases, fundamental mathematical tools and algorithms related to centrality measures and network matrix properties are actively being developed. Indeed, the PageRank algorithm, which was initially at the basis of the development of the Google search engine (Brin and Page , 1998; Langville and Meyer, 2006), is directly linked to the mathematical properties of Markov chains (Markov , 1906) and Perron-Frobenius operators (Brin and Stuck, 2002; Langville and Meyer, 2006). Due to its mathematical foundation, this algorithm determines a ranking order of nodes that can be applied to various types of directed networks. However, the recent enormous development of WWW and communication networks requires the creation of new tools and algorithms to characterize the properties of these networks on a more detailed and precise level. For example, such networks contain weakly coupled or secret communities which may correspond to very small values of the PageRank and are hard to detect. It is therefore highly important to have new methods to classify and rank enormous amount of network information in a way adapted to internal network structures and characteristics.

This review describes matrix tools and algorithms which facilitate classification and information retrieval from large networks recently created by human activity. The Google matrix formed by links of the network has typically a huge size. Thus, the analysis of its spectral properties including complex eigenvalues and eigenvectors represents a challenge for analytical and numerical methods. It is rather surprising, but the class of such matrices, belonging to the class of Markov chains and Perron-Frobenius operators, was practically not investigated in physics. Indeed, usually the physical problems belong to the class of Hermitian or unitary matrices. Their properties had been actively studied in the frame of Random Matrix Theory (RMT) (Akemann *et al.*, 2011; Guhr *et al.*, 1998; Mehta, 2004) and quantum chaos (Haake, 2010). The analytical and numerical tools developed in these research fields allowed to understand many universal and peculiar features of such matrices in the limit of large matrix size corresponding to many-body quantum systems (Guhr *et al.*, 1998), quantum computers (Shepelyansky , 2001) and a semiclassical limit of large quantum numbers in the regime of quantum chaos (Haake, 2010). In contrast to the Hermitian problem, the Google matrices of directed networks have complex eigenvalues. The only physical systems where similar matrices had been studied analytically and numerically correspond to models of quantum chaotic scattering whose spectrum is known to have such unusual properties as the fractal Weyl law (Gaspard, 2014; Nonnenmacher and Zworski , 2007; Shepelyansky , 2008; Sjöstrand , 1990; Zworski , 1999).

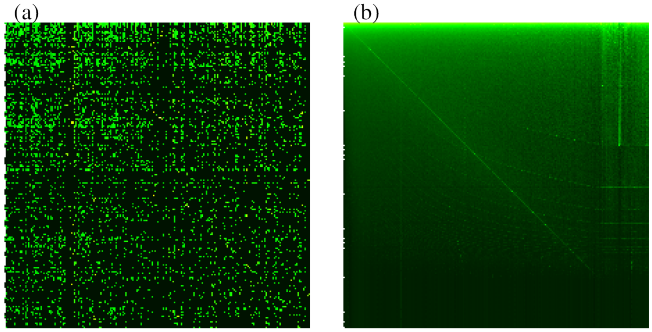

FIG. 1 (Color online) Google matrix of the network Wikipedia English articles for Aug 2009 in the basis of PageRank index K (and K'). Matrix $G_{KK'}$ corresponds to x (and y) axis with $1 \leq K, K' \leq 200$ on panel (a), and with $1 \leq K, K' \leq N$ on panel (b); all nodes are ordered by PageRank index K of matrix G and thus we have two matrix indexes K, K' for matrix elements in this basis. Panel (a) shows the first 200×200 matrix elements of G matrix (see Sec. III). Panel (b) shows density of all matrix elements coarse-grained on 500×500 cells where its elements, $G_{K,K'}$, are written in the PageRank basis $K(i)$ with indexes $i \to K(i)$ (in x-axis) and $j \to K'(j)$ (in a usual matrix representation with $K = K' = 1$ on the top-left corner). Color shows the density of matrix elements changing from black for minimum value $((1-\alpha)/N)$ to white for maximum value via green (gray) and yellow (light gray); here the damping factor is $\alpha = 0.85$ After (Ermann *et al.*, 2012a).

In this review we present extensive analysis of a variety of Google matrices emerging from real networks in various sciences including WWW of UK universities, Wikipedia, Physical Review citation network, Linux Kernel network, world trade network from the UN COMTRADE database, brain neural networks, networks of DNA sequences and many others. As an example, the Google matrix of Wikipedia network of English articles (2009) is shown in Fig. 1. We demonstrate that the analysis of the spectrum and eigenstates of a Google matrix of a given network provides a detailed understanding about the information flow and ranking. We also show that such type of matrices naturally appear for Ulam networks of dynamical maps (Frahm and Shepelyansky , 2012b; She-

pelyansky and Zhirov , 2010a) in the framework of the Ulam method (Ulam, 1960).

At present, Wikipedia, a free online encyclopaedia, stores more and more information becoming the largest database of human knowledge. In this respect it is similar to the Library of Babel, described by Jorge Luis Borges (Borges, 1962). The understanding of hidden relations between various areas of knowledge on the basis of Wikipedia can be improved with the help of Google matrix analysis of directed hyperlink network of Wikipedia articles as described in this review.

The RMT and quantum chaos tools, combined with the efficient numerical methods for large matrix diagonalization like the Arnoldi method (Stewart, 2001), allow to analyze the spectral properties of such large matrices as an entire Twitter network of 41 millions users (Frahm and Shepelyansky , 2012b). In 1998 Brin and Page pointed out that *"despite the importance of large-scale search engines on the web, very little academic research has been done on them"* (Brin and Page , 1998). We hope that this review provides solid mathematical basis of matrix methods of efficient analysis of directed networks emerging in various sciences. The described methods will find broad interdisciplinary applications in mathematics, physics and computer science with the cross-fertilization of different research fields.

An interested reader can find a general information about complex networks (see also Sec. II) in well established papers, reviews and books (Watts and Strogatz , 1998), (Albert and Barabási , 2002; Caldarelli, 2003; Newman , 2003), (Castellano *et al.*, 2009; Dorogovtsev *et al.*, 2008), (Dorogovtsev, 2010; Fortunato , 2010; Newman, 2010). Descriptions of Markov chains and Perron-Frobenius operators are given in (Brin and Page , 1998; Langville and Meyer, 2006) while properties of Random Matrix Theory (RMT) and quantum chaos are described in (Akemann *et al.*, 2011; Guhr *et al.*, 1998; Haake, 2010; Mehta, 2004).

The data sets of the main part of networks considered here are available at (FETNADINE database, 2014) from Quantware group.

II. SCALE-FREE PROPERTIES OF DIRECTED NETWORKS

The distributions of the number of ingoing or outgoing links per node for directed networks with N nodes and N_ℓ links are well known as indegree and outdegree distributions in the community of computer science (Caldarelli, 2003; Donato *et al.*, 2004; Pandurangan *et al.*, 2005). A network is described by an adjacency matrix A_{ij} of size $N \times N$ with $A_{ij} = 1$ when there is a link from a node j to a node i in the network, i. e. "j points to i", and $A_{ij} = 0$ otherwise. Real networks are often characterized by power law distributions for the number of ingoing and outgoing links per node $w_{\rm in,out}(k) \propto 1/k^{\mu_{\rm in,out}}$ with typical exponents $\mu_{\rm in} \approx 2.1$ and $\mu_{\rm out} \approx 2.7$ for the WWW. For example, for the Wikipedia network of Fig. 1 one finds $\mu_{\rm in} = 2.09 \pm 0.04$, $\mu_{\rm out} = 2.76 \pm 0.06$ as shown in Fig. 2 (Zhirov *et al.*, 2010).

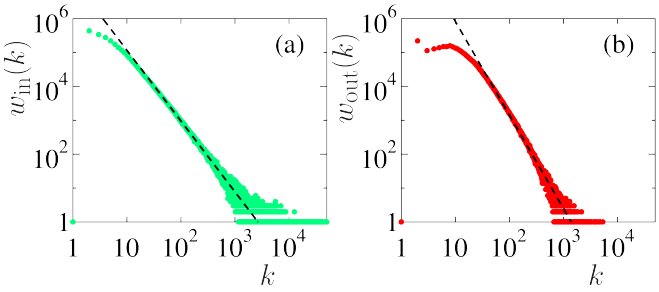

FIG. 2 (Color online) Distribution $w_{\rm in,out}(k)$ of number of ingoing (a) and outgoing (b) links k for $N = 3282257$ Wikipedia English articles (Aug 2009) of Fig. 1 with total number of links $N_\ell = 71012307$. The straight dashed fit line shows the slope with $\mu_{\rm in} = 2.09 \pm 0.04$ (a) and $\mu_{\rm out} = 2.76 \pm 0.06$ (b). After (Zhirov *et al.*, 2010).

Statistical preferential attachment models were initially developed for undirected networks (Albert and Barabási , 2000). Their generalization to directed networks (Giraud *et al.*, 2009) generates a power law distribution for ingoing links with $\mu_{\rm in} \approx 2$ but the distribution of outgoing links is more close to an exponential decay. We will see below that these models are not able to reproduce the spectral properties of G in real networks.

The most recent studies of WWW, crawled by the Common Crawl Foundation in 2012 (Meusel *et al.*, 2014) for $N \approx 3.5 \times 10^9$ nodes and $N_\ell \approx 1.29 \times 10^{11}$ links, provide the exponents $\mu_{\rm in} \approx 2.24$, $\mu_{\rm out} \approx 2.77$, even if the authors stress that these distributions describe probabilities at the tails which capture only about one percent of nodes. Thus, at present the existing statistical models of networks capture only in an approximate manner the real situation in large networks.

III. CONSTRUCTION OF GOOGLE MATRIX AND ITS PROPERTIES

A. Construction rules

The matrix S_{ij} of Markov transitions (Markov , 1906) is constructed from the adjacency matrix $A_{ij} \to S_{ij}$ by normalizing elements of each column so that their sum is equal to unity ($\sum_i S_{ij} = 1$) and replacing columns with only zero elements (*dangling nodes*) by $1/N$. Such matrices with columns sum normalized to unity and $S_{ij} \geq 0$ belong to the class of Perron-Frobenius operators with a possibly degenerate unit eigenvalue $\lambda = 1$ and other eigenvalues obeying $|\lambda| \leq 1$ (see Sec. III.B). Then the Google matrix of the network is introduced as: (Brin and Page , 1998)

$$G_{ij} = \alpha S_{ij} + (1 - \alpha)/N \ . \qquad (1)$$

The damping factor α in the WWW context describes the probability $(1-\alpha)$ to jump to any node for a random surfer. For WWW the Google search engine uses $\alpha \approx 0.85$ (Langville and Meyer, 2006). For $0 \leq \alpha \leq 1$ the matrix G also belongs to the class of Perron-Frobenius operators as S and with its columns sum normalized. However, for $\alpha < 1$ its largest eigenvalue $\lambda = 1$ is not degenerate and the other eigenvalues lie inside a smaller circle of radius α, i.e. $|\lambda| \leq \alpha$ (Brin and Stuck, 2002; Langville and Meyer, 2006).

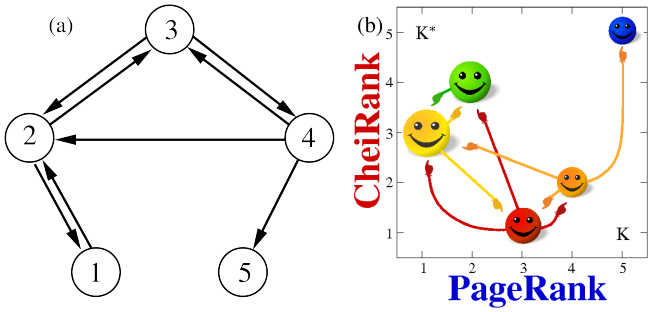

FIG. 3 (Color online) (a) Example of simple network with directed links between 5 nodes. (b) Distribution of 5 nodes from (a) on the PageRank-CheiRank plane (K, K^*), where the size of node is proportional to PageRank probability $P(K)$ and color of node is proportional to CheiRank probability $P^*(K^*)$, with maximum at red/gray and minimum at blue/black; the location of nodes of panel (a) on $(K_i, K_i{}^*)$ plane is: $(2,4)$, $(1,3)$, $(3,1)$, $(4,2)$, $(5,5)$ for original nodes $i = 1,2,3,4,5$ respectively; PageRank and CheiRank vectors are computed from the Google matrices G and G^* shown in Fig. 4 at a damping factor $\alpha = 0.85$.

The right eigenvector at $\lambda = 1$, which is called the PageRank, has real nonnegative elements $P(i)$ and gives the probability $P(i)$ to find a random surfer at site i. The PageRank can be efficiently determined by the power iteration method which consists of repeatedly multiplying G to an iteration vector which is initially chosen as a given random or uniform initial vector. Developing the initial vector in a basis of eigenvectors of G one finds that the other eigenvector coefficients decay as $\sim \lambda^n$ and only the PageRank component, with $\lambda = 1$, survives in the limit $n \to \infty$. The finite gap $1 - \alpha \approx 0.15$ between the largest eigenvalue and other eigenvalues ensures, after several tens of iterations, the fast exponential convergence of the method also called the "PageRank algorithm". A multiplication of G to a vector requires only $O(N_\ell)$ multiplications due to the links and the additional contributions due to dangling nodes and damping factor can be efficiently performed with $O(N)$ operations. Since often the average number of links per node is of the order of a few tens for WWW and many other networks one has effectively N_ℓ and N of the same order of magnitude. At $\alpha = 1$ the matrix G coincides with the matrix S and we will see below in Sec. VIII that for this case the largest eigenvalue $\lambda = 1$ is usually highly degenerate due to many invariant subspaces which define many independent Perron-Frobenius operators with at least one eigenvalue $\lambda = 1$ for each of them.

Once the PageRank is found, e.g. at $\alpha = 0.85$, all nodes can be sorted by decreasing probabilities $P(i)$. The node rank is then given by the index $K(i)$ which reflects the relevance of the node i. The top PageRank nodes, with largest probabilities, are located at small values of $K(i) = 1, 2, ...$

It is known that the PageRank probability is proportional to the number of ingoing links (Langville and Meyer, 2006; Litvak *et al.*, 2008), characterizing how popular or known a given node is. Assuming that the PageRank probability decays algebraically as $P_i \sim 1/K_i^\beta$ we obtain that the number of nodes N_P with PageRank probability P scales as $N_P \sim 1/P^{\mu_{\rm in}}$ with $\mu_{\rm in} = 1 + 1/\beta$ so that $\beta \approx 0.9$ for $\mu_{\rm in} \approx 2.1$ being in a agreement with the numerical data for WWW (Donato *et al.*, 2004; Meusel *et al.*, 2014; Pandurangan *et al.*, 2005) and Wikipedia network (Zhirov *et al.*, 2010).

In addition to a given directed network with adjacency matrix A it is useful to analyze an inverse network where links are inverted and whose adjacency matrix A^* is the transpose of A, i.e. $A^*_{ij} = A_{ji}$. The matrices S^* and the Google matrix G^* of the inverse network are then constructed in the same way from A^* as described above and according to the relation (1) using the same value of α as for the G matrix. The right eigenvector of G^* at eigenvalue $\lambda = 1$ is called CheiRank giving a complementary rank index $K^*(i)$ of network nodes (Chepelianskii, 2010; Ermann *et al.*, 2012a; Zhirov *et al.*, 2010). The CheiRank probability $P^*(K^*)$ is proportional to the number of outgoing links highlighting node communicativity (see e.g. (Ermann *et al.*, 2012a; Zhirov *et al.*, 2010)). In analogy with the PageRank we obtain that $P^* \sim 1/K^{*\beta}$ with $\beta = 1/(\mu_{\rm out} - 1) \approx 0.6$ for typical $\mu_{\rm out} \approx 2.7$. The statistical properties of distribution of nodes on the PageRank-CheiRank plane are described in (Ermann *et al.*, 2012a) for various directed networks. We will discuss them below.

(a) $A = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$

(b) $A^* = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$

(c) $S = \begin{pmatrix} 0 & 1/2 & 1/3 & 0 & 1/5 \\ 1 & 0 & 1/3 & 1/3 & 1/5 \\ 0 & 1/2 & 0 & 1/3 & 1/5 \\ 0 & 0 & 1/3 & 0 & 1/5 \\ 0 & 0 & 0 & 1/3 & 1/5 \end{pmatrix}$

(d) $S^* = \begin{pmatrix} 0 & 1/3 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 1/2 & 1/3 & 0 & 1 & 0 \\ 0 & 1/3 & 1/2 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$

(e) $G = \begin{pmatrix} 0.03 & 0.455 & 0.313 & 0.03 & 0.2 \\ 0.88 & 0.03 & 0.313 & 0.313 & 0.2 \\ 0.03 & 0.455 & 0.03 & 0.313 & 0.2 \\ 0.03 & 0.03 & 0.313 & 0.03 & 0.2 \\ 0.03 & 0.03 & 0.03 & 0.313 & 0.2 \end{pmatrix}$

(f) $G^* = \begin{pmatrix} 0.03 & 0.313 & 0.03 & 0.03 & 0.03 \\ 0.455 & 0.03 & 0.455 & 0.03 & 0.03 \\ 0.455 & 0.313 & 0.03 & 0.88 & 0.03 \\ 0.03 & 0.313 & 0.455 & 0.03 & 0.88 \\ 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \end{pmatrix}$

FIG. 4 (a) Adjacency matrix A of network of Fig. 3(a) with indexes used there, (b) adjacency matrix A^* for the network with inverted links; matrices S (c) and S^* (d) corresponding to the matrices A, A^*; the Google matrices G (e) and G^* (f) corresponding to matrices S and S^* for $\alpha = 0.85$ (only 3 digits of matrix elements are shown).

For an illustration we consider an example of a simple network of five nodes shown in Fig. 3(a). The corresponding adjacency matrices A, A^* are shown in Fig. 4 for the indexes given in Fig. 3(a). The matrices of Markov transitions S, S^* and Google matrices are computed as described above and from Eq. (1). The distribution of nodes on (K, K^*) plane is shown in Fig. 3(b). After permutations the matrix G can be rewritten in the basis of PageRank index K as it is done in Fig. 1.

B. Markov chains and Perron-Frobenius operators

Matrices with real non-negative elements and column sums normalized to unity belong to the class of Markov chains (Markov , 1906) and Perron-Frobenius operators (Brin and Stuck, 2002), which have been used in a mathematical analysis of dynamical systems. A numerical analysis of finite size approximants of such operators is closely linked with the Ulam method (Ulam, 1960) which naturally generates such matrices for dynamical maps (Ermann and Shepelyansky , 2010a,b; Shepelyansky and Zhirov , 2010a). The Ulam method generates Ulam networks whose properties are discussed in Sec.VI.

Matrices G of this type have at least (one) unit eigenvalue $\lambda = 1$ since the vector $e^T = (1, \ldots, 1)$ is obviously a left eigenvector for this eigenvalue. Furthermore one verifies easily that for any vector v the inequality $\|G\, v\|_1 \leq \|v\|_1$ holds where the norm is the standard 1-norm. From this inequality one obtains immediately that all eigenvalues λ of G lie in a circle of radius unity: $|\lambda| \leq 1$. For the Google matrix G as given in (1) one can furthermore show for $\alpha < 1$ that the unity eigenvalue is not degenerate and the other eigenvalues obey even $|\lambda| \leq \alpha$ (Langville and Meyer, 2006).

It should be pointed out that due to the asymmetry of links on directed networks such matrices have in general a complex eigenvalue spectrum and sometimes they are not even diagonalizable, i.e. there may also be generalized eigenvectors associated to non-trivial Jordan blocks. Matrices of this type rarely appear in physical problems which are usually characterized by Hermitian or unitary matrices with real eigenvalues or located on the unitary circle. The universal spectral properties of such hermitian or unitary matrices are well described by RMT (Akemann *et al.*, 2011; Guhr *et al.*, 1998; Haake, 2010). In contrast to this non-trivial complex spectra appear in physical systems only in problems of quantum chaotic scattering and systems with absorption. In such cases it may happen that the number of states N_γ, with finite values $0 < \lambda_{\min} \leq |\lambda| \leq 1$ ($\gamma = -2 \ln |\lambda|$), can grow algebraically $N_\gamma \propto N^\nu$ with increasing matrix size N, with an exponent $\nu < 1$ corresponding to a fractal Weyl law proposed first in mathematics (Sjöstrand , 1990). Therefore most of eigenvalues drop to $\lambda = 0$ with $N \to \infty$. We discuss this unusual property in Sec.V.

C. Invariant subspaces

For typical networks the set of nodes can be decomposed in invariant *subspace nodes* and fully connected *core space nodes* leading to a block structure of the matrix S in (1) which can be represented as (Frahm *et al.*, 2011):

$$S = \begin{pmatrix} S_{ss} & S_{sc} \\ 0 & S_{cc} \end{pmatrix} . \qquad (2)$$

The core space block S_{cc} contains the links between core space nodes and the coupling block S_{sc} may contain links from certain core space nodes to certain invariant subspace nodes. By construction there are no links from nodes of invariant subspaces to the nodes of core space. Thus the subspace-subspace block S_{ss} is actually composed of many diagonal blocks for many invariant subspaces whose number can generally be rather large. Each of these blocks corresponds to a column sum normalized matrix with positive elements of the same type as G and has therefore at least one unit eigenvalue. This leads to a high degeneracy N_1 of the eigenvalue $\lambda = 1$ of S, for example $N_1 \sim 10^3$ as for the case of UK universities (see Sec. VIII).

In order to obtain the invariant subspaces, we determine iteratively for each node the set of nodes that can be reached by a chain of non-zero matrix elements of S. If this set contains all nodes (or at least a macroscopic fraction) of the network, the initial node belongs to the *core space* V_c. Otherwise, the limit set defines a subspace which is invariant with respect to applications of the matrix S. At a second step all subspaces with common members are merged resulting in a sequence of disjoint subspaces V_j of dimension d_j and which are invariant by applications of S. This scheme, which can be efficiently implemented in a computer program, provides a subdivision over N_c core space nodes (70-80% of N for UK university networks) and $N_s = N - N_c$ subspace nodes belonging to at least one of the invariant subspaces V_j. This procedure generates the block triangular structure (2). One may note that since a dangling node is connected by construction to all other nodes it belongs obviously to the core space as well as all nodes which are linked (directly or indirectly) to a dangling node. As a consequence the invariant subspaces do not contain dangling nodes nor nodes linked to dangling nodes.

The detailed algorithm for an efficient computation of the invariant subspaces is described in (Frahm *et al.*, 2011). As a result the total number of all subspace nodes N_s, the number of independent subspaces N_d, the maximal subspace dimension $d_{\max}$ etc. can be determined. The statistical properties for the distribution of subspace dimensions are discussed in Sec. VIII for UK universities and Wikipedia networks. Furthermore it is possible to determine numerically with a very low effort the eigenvalues of S associated to each subspace by separate diagonalization of the corresponding diagonal blocks in the matrix S_{ss}. For this, either exact diagonalization or, in

rare cases of quite large subspaces, the Arnoldi method (see the next subsection) can be used.

After the subspace eigenvalues are determined one can use the Arnoldi method to the projected core space matrix block S_{cc} to determine the leading core space eigenvalues. In this way one obtains accurate eigenvalues because the Arnoldi method does not need to compute the numerically very problematic highly degenerate unit eigenvalues of S since the latter are already obtained from the separate and cheap subspace diagonalization. Actually the alternative and naive application of the Arnoldi method on the full matrix S, without computing the subspaces first, does not provide the correct number N_1 of degenerate unit eigenvalues and also the obtained clustered eigenvalues, close to unity, are not very accurate. Similar problems hold for the full matrix G (with damping factor $\alpha < 1$) since here only the first eigenvector, the PageRank, can be determined accurately but there are still many degenerate (or clustered) eigenvalues at (or close to) $\lambda = \alpha$.

Since the columns sums of S_{cc} are less than unity, due to non-zero matrix elements in the block S_{sc}, the leading core space eigenvalue of S_{cc} is also below unity $|\lambda_1^{(\mathrm{core})}| < 1$ even though in certain cases the gap to unity may be very small (see Sec. VIII).

We consider concrete examples of such decompositions in Sec. VIII and show in this review spectra with subspace and core space eigenvalues of matrices S for several network examples. The mathematical results for properties of the matrix S are discussed in (Serra-Capizzano , 2005).

D. Arnoldi method for numerical diagonalization

The most adapted numerical method to determine the largest eigenvalues of large sparse matrices is the Arnoldi method (Arnoldi , 1951; Frahm and Shepelyansky , 2010; Golub and Greif , 2006; Stewart, 2001). Indeed, usually the matrix S in Eq. (1) is very sparse with only a few tens of links per node $\zeta = N_\ell/N \sim 10$. Thus, a multiplication of a vector by G or S is numerically cheap. The Arnoldi method is similar in spirit to the Lanzcos method, but is adapted to non-Hermitian or non-symmetric matrices. Its main idea is to determine recursively an orthonormal set of vectors $\xi_0, \ldots \xi_{n_\mathrm{A}-1}$, which define a *Krylov space*, by orthogonalizing $S\xi_k$ on the previous vectors $\xi_0, \ldots \xi_k$ by the Gram-Schmidt procedure to obtain ξ_{k+1} and where ξ_0 is some normalized initial vector. The dimension n_A of the Krylov space (in the following called the *Arnoldi-dimension*) should be "modest" but not too small. During the Gram-Schmidt procedure one obtains furthermore the explicit expression: $S\xi_k = \sum_{j=0}^{k+1} h_{jk}\,\xi_j$ with matrix elements h_{jk}, of the Arnoldi representation matrix of S on the Krylov space, given by the scalar products or inverse normalization constants calculated during the orthogonalization. In order to obtain a closed representation matrix one needs to replace the last coupling element $h_{n_\mathrm{A},n_\mathrm{A}-1} \to 0$ which introduces a mathematical approximation. The eigenvalues of the $n_\mathrm{A} \times n_\mathrm{A}$ matrix h are called the *Ritz eigenvalues* and represent often very accurate approximations of the exact eigenvalues of S, at least for a considerable fraction of the Ritz eigenvalues with largest modulus.

In certain particular cases, when ξ_0 belongs to an S invariant subspace of small dimension d, the element $h_{d,d-1}$ vanishes automatically (if $d \leq n_\mathrm{A}$ and assuming that numerical rounding errors are not important) and the Arnoldi iteration stops at $k = d$ and provides d exact eigenvalues of S for the invariant subspace. One can mention that there are more sophisticated variants of the Arnoldi method (Stewart, 2001) where one applies (implicit) modifications on the initial vector ξ_0 in order to force this vector to be in some small dimensional invariant subspace which results in such a vanishing coupling matrix element. These variants known as (implicitly) restarted Arnoldi methods allow to concentrate on certain regions on the complex plane to determine a few but very accurate eigenvalues in these regions. However, for the cases of Google matrices, where one is typically interested in the largest eigenvalues close to the unit circle, only the basic variant described above was used but choosing larger values of n_A as would have been possible with the restarted variants. The initial vector was typically chosen to be random or as the vector with unit entries.

Concerning the numerical resources the Arnoldi method requires ζN double precision registers to store the non-zero matrix elements of S, $n_\mathrm{A} N$ registers to store the vectors ξ_k and const.$\times n_\mathrm{A}^2$ registers to store h (and various copies of h). The computational time scales as $\zeta\, n_\mathrm{A}\, N_d$ for the computation of $S\,\xi_k$, with $N_d\, n_\mathrm{A}^2$ for the Gram-Schmidt orthogonalization procedure (which is typically dominant) and with const.$\times n_\mathrm{A}^3$ for the diagonalization of h.

The details of the Arnoldi method are described in Refs. given above. This method has problems with degenerate or strongly clustered eigenvalues and therefore for typical examples of Google matrices it is applied to the core space block S_{cc} where the effects of the invariant subspaces, being responsible for most of the degeneracies, are exactly taken out according to the discussion of the previous subsection. In typical examples it is possible to find about $n_\mathrm{A} \approx 640$ eigenvalues with largest $|\lambda|$ for the entire Twitter network with $N \approx 4.1 \times 10^7$ (see Sec. X) and about $n_\mathrm{A} \approx 6000$ eigenvalues for Wikipedia networks with $N \approx 3.2 \times 10^6$ (see Sec. IX). For the two university networks of Cambridge and Oxford 2006 with $N \approx 2 \times 10^5$ it is possible to compute $n_\mathrm{A} \approx 20000$ eigenvalues (see Sec. VIII). For the case of the Citation network of Physical Review (see Sec. XII) with $N \approx 4.6 \times 10^5$ it is even possible and necessary to use high precision computations (with up to 768 binary digits) to determine accurately the Arnoldi matrix h with $n_\mathrm{A} \approx 2000$ (Frahm *et al.*, 2014b).

E. General properties of eigenvalues and eigenstates

According to the Perron-Frobenius theorem all eigenvalues λ_i of G are distributed inside the unitary circle $|\lambda| \leq 1$. It can be shown that at $\alpha < 1$ there is only one eigenvalue $\lambda_0 = 1$ and all other $|\lambda_i| \leq \alpha$ having a simple dependence on α: $\lambda_i \to \alpha\lambda_i$ (see e.g. (Langville and Meyer, 2006)). The right eigenvectors $\psi_i(j)$ are defined by the equation

$$\sum_{j'} G_{jj'}\psi_i(j') = \lambda_i\psi_i(j) \ . \qquad (3)$$

Only the PageRank vector is affected by α while other eigenstates are independent of α due to their orthogonality to the left unit eigenvector at $\lambda = 1$. Left eigenvectors are orthonormal to right eigenvectors (Langville and Meyer, 2006).

It is useful to characterize the eigenvectors by their Inverse Participation Ratio (IPR) $\xi_i = (\sum_j |\psi_i(j)|^2)^2 / \sum_j |\psi_i(j)|^4$ which gives an effective number of nodes populated by an eigenvector ψ_i. This characteristics is broadly used for description of localized or delocalized eigenstates of electrons in a disordered potential with Anderson transition (see e.g. (Evers and Mirlin , 2008; Guhr *et al.*, 1998)). We discuss the specific properties of eigenvectors in next Secs.

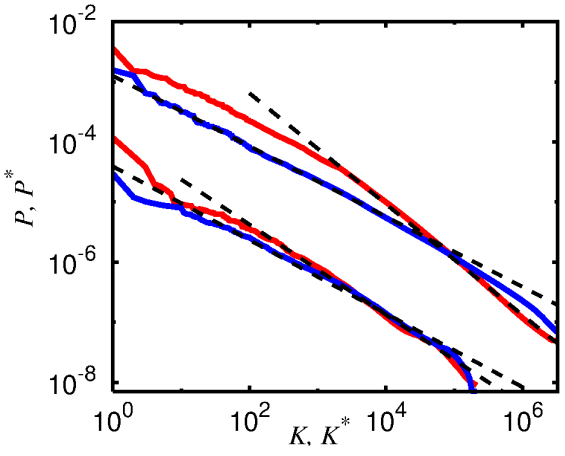

FIG. 5 (Color online) Dependence of probabilities of PageRank P (red/gray curve) and CheiRank P^* (blue/black curve) vectors on the corresponding rank indexes K and K^* for networks of Wikipedia Aug 2009 (top curves) and University of Cambridge (bottom curves, moved down by a factor 100). The straight dashed lines show the power law fits for PageRank and CheiRank with the slopes $\beta = 0.92; 0.58$ respectively, corresponding to $\beta = 1/(\mu_{\rm in,out} - 1)$ for Wikipedia (see Fig. 2), and $\beta = 0.75, 0.61$ for Cambridge. After (Zhirov *et al.*, 2010) and (Frahm *et al.*, 2011).

IV. CHEIRANK VERSUS PAGERANK

It is established that ranking of network nodes based on PageRank order works reliably not only for WWW but also for other directed networks. As an example it is possible to quote the citation network of Physical Review (Radicchi *et al.*, 2009; Redner , 1998, 2005), Wikipedia network (Aragón *et al.*, 2012; Eom and Shepelyansky , 2013a; Skiena and Ward, 2014; Zhirov *et al.*, 2010) and even the network of world commercial trade (Ermann and Shepelyansky , 2011b). Here we describe the main properties of PageRank and CheiRank probabilities using a few real networks. More detailed presentation for concrete networks follows in next Secs.

A. Probability decay of PageRank and CheiRank

Wikipedia is a useful example of a scale-free network. An article quotes other Wikipedia articles that generates a network of directed links. For Wikipedia of English articles dated by Aug 2009 we have $N = 3282257$, $N_\ell = 71012307$ ((Zhirov *et al.*, 2010)). The dependencies of PageRank $P(K)$ and CheiRank $P^*(K^*)$ probabilities on indexes K and K^* are shown in Fig. 5. In a large range the decay can be satisfactory described by an algebraic law with an exponent β. The obtained β values are in a reasonable agreement with the expected relation $\beta = 1/(\mu_{\rm in,out} - 1)$ with the exponents of distribution of links given above. However, the decay is algebraic only on a tail, showing certain nonlinear variations well visible for $P^*(K^*)$ at large values of P^*.

Similar data for network of University of Cambridge (2006) with $N = 212710$, $N_\ell = 2015265$ (Frahm *et al.*, 2011) are shown in the same Fig. 5. Here, the exponents β have different values with approximately the same statistical accuracy of β.

Thus we come to the same conclusion as (Meusel *et al.*, 2014): the probability decay of PageRank and CheiRank is only approximately algebraic, the relation between exponents β and μ also works only approximately.

B. Correlator between PageRank and CheiRank

Each network node i has both PageRank $K(i)$ and CheiRank $K(i)^*$ indexes so that it is interesting to know what is a correlation between the corresponding vectors of PageRank and CheiRank. It is convenient to characterized this by a correlator introduced in (Chepelianskii, 2010)

$$\kappa = N \sum_{i=1}^{N} P(K(i))P^*(K^*(i)) - 1 \ . \qquad (4)$$

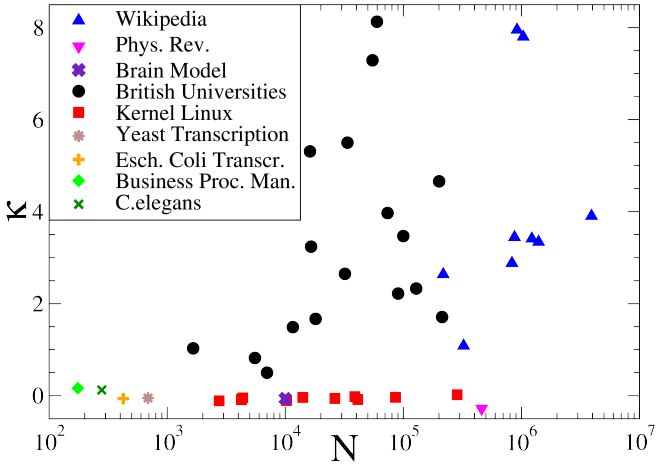

FIG. 6 (Color online) Correlator κ as a function of the number of nodes N for different networks: Wikipedia networks, Phys Rev network, 17 UK universities, 10 versions of Kernel Linux Kernel PCN, Escherichia Coli and Yeast Transcription Gene networks, Brain Model Network, C.elegans neural network and Business Process Management Network. After (Ermann *et al.*, 2012a) with additional data from (Abel and Shepelyansky , 2011), (Eom and Shepelyansky , 2013a), (Kandiah and Shepelyansky , 2014a), (Frahm *et al.*, 2014b).

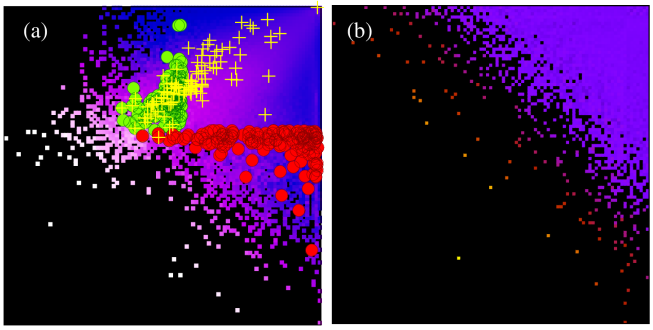

FIG. 7 (Color online) Density distribution of network nodes $W(K,K^*) = dN_i/dKdK^*$ shown on the plane of PageRank and CheiRank indexes in logscale $(\log_N K, \log_N K^*)$ for all $1 \le K, K^* \le N$, density is computed over equidistant grid in plane $(\log_N K, \log_N K^*)$ with 100×100 cells; color shows average value of W in each cell, the normalization condition is $\sum_{K,K^*} W(K,K^*) = 1$. Density $W(K,K^*)$ is shown by color with blue (dark gray) for minimum in (a),(b) and white (a) and yellow (white) (b) for maximum (black for zero). Panel (a): data for Wikipedia Aug (2009), $N = 3282257$, green/red (light gray/dark gray) points show top 100 persons from PageRank/CheiRank, yellow (white) pluses show top 100 persons from (Hart, 1992); after (Zhirov *et al.*, 2010). Panel (b): Density distribution $W(K,K^*) = dN_i/dKdK^*$ for Linux Kernel V2.4 network with $N = 85757$, after (Ermann *et al.*, 2012a).

Even if all the networks from Fig. 6 have similar algebraic decay of PageRank probability with K and similar $\beta \sim 1$ exponents we see that the correlations between PageRank and CheiRank vectors are drastically different in these networks. Thus the networks of UK universities and 9 different language editions of Wikipedia have the correlator $\kappa \sim 1 - 8$ while all other networks have $\kappa \sim 0$. This means that there are significant differences hidden in the network architecture which are no visible from PageRank analysis. We will discuss the possible origins of such a difference for the above networks in next Secs.

C. PageRank-CheiRank plane

A more detailed characterization of correlations between PageRank and CheiRank vectors can be obtained from a distribution of network nodes on the two-dimensional plane (2D) of indexes (K, K^*). Two examples for Wikipedia and Linux networks are shown in Fig. 7. A qualitative difference between two networks is obvious. For Wikipedia we have a maximum of density along the line $\ln K^* \approx 5 + (\ln K)/3$ that results from a strong correlation between PageRank and CheiRank with $\kappa = 4.08$. In contrast to that for the Linux network V2.4 we have a homogeneous density distribution of nodes along lines $\ln K^* = \ln K + const$ corresponding to uncorrelated probabilities $P(K)$ and $P^*(K^*)$ and even slightly negative value of $\kappa = -0.034$. We note that if for Wikipedia we generate nodes with independent probabilities distributions P and P^*, obtained from this network at the corresponding value of N, then we obtain a homogeneous node distribution in (K, K^*) plane (in $(\log K, \log K^*)$ plane it takes a triangular form, see Fig.4 at (Zhirov *et al.*, 2010)).

In Fig. 7(a) we also show the distribution of top 100 persons from PageRank and CheiRank compared with the top 100 persons from (Hart, 1992). There is a significant overlap between PageRank and Hart ranking of persons while CheiRank generates mainly another listing of people. We discuss the Wikipedia ranking of historical figures in Sec. IX.

D. 2DRank

PageRank and CheiRank indexes $K_i K_i^*$ order all network nodes according to a monotonous decrease of corresponding probabilities $P(K_i)$ and $P^*(K_i^*)$. While top K nodes are most popular or known in the network, top K^* nodes are most communicative nodes with many outgoing links. It is useful to consider an additional ranking K_2, called 2DRank, which combines properties of both ranks K and K^* (Zhirov *et al.*, 2010).

The ranking list $K_2(i)$ is constructed by increasing $K \to K+1$ and increasing 2DRank index $K_2(i)$ by one if a new entry is present in the list of first $K^* < K$ entries of CheiRank, then the one unit step is done in K^* and K_2 is increased by one if the new entry is present in the list of first $K < K^*$ entries of CheiRank. More

formally, 2DRank $K_2(i)$ gives the ordering of the sequence of sites, that appear inside the squares $[1,1;\ K=k, K^*=k;\ ...]$ when one runs progressively from $k=1$ to N. In fact, at each step $k \to k+1$ there are tree possibilities: (i) no new sites on two edges of square, (ii) only one site is on these two edges and it is added in the listing of $K_2(i)$ and (iii) two sites are on the edges and both are added in the listing $K_2(i)$, first with $K > K^*$ and second with $K < K^*$. For (iii) the choice of order of addition in the list $K_2(i)$ affects only some pairs of neighboring sites and does not change the main structure of ordering. An illustration example of 2DRank algorithm is given in Fig.7 at (Zhirov *et al.*, 2010). For Wikipedia 2DRanking of persons is discussed in Sec. IX.

E. Historical notes on spectral ranking

Starting from the work of Markov (Markov , 1906) many scientists contributed to the development of spectral ranking of Markov chains. Research of Perron (1907) and Frobenius (1912) led to the Perron-Frobenius theorem for square matrices with positive entries (see e.g. (Brin and Stuck, 2002)). Important steps have been done by researchers in psychology, sociology and mathematics including J.R.Seeley (1949), T.-H.Wei (1952), L.Katz (1953), C.H.Hubbell (1965). The detailed historical description of spectral ranking research is reviewed by (Franceschet , 2011) and (Vigna, 2013). In the WWW context, the Google matrix in the form (1), with regularization of dangling nodes and damping factor α, was introduced by (Brin and Page , 1998).

A PageRank vector of a Google matrix G^* with inverted directions of links has been considered by (Fogaras , 2003) and (Hrisitidis *et al.*, 2008), but no systematic statistical analysis of 2DRanking was presented there. An important step was done by (Chepelianskii, 2010) who analyzed $\lambda = 1$ eigenvectors of G for directed network and of G^* for network with inverted links. The comparative analysis of Linux Kernel network and WWW of University of Cambridge demonstrated a significant differences in correlator κ values on these networks and different functions of top nodes in K and K^*. The term CheiRank was coined in (Zhirov *et al.*, 2010) to have a clear distinction between eigenvectors of G and G^*. We note that top PageRank and CheiRank nodes have certain similarities with authorities and hubs appearing in the HITS algorithm (Kleinberg , 1999). However, the HITS is query dependent while the rank probabilities $P(K_i)$ and $P^*({K_i}^*)$ classify all nodes of the network.

V. COMPLEX SPECTRUM AND FRACTAL WEYL LAW

The Weyl law (Weyl , 1912) gives a fundamental link between the properties of quantum eigenvalues in closed Hamiltonian systems, the Planck constant $\hbar$ and the classical phase space volume. The number of states in this case is determined by the phase volume of a system with dimension d. The case of Hermitian operators is now well understood both on mathematical and physical grounds (Dimassi and Sjöstrand, 1999; Landau and Lifshitz, 1989). Surprisingly, only recently it has been realized that the case of nonunitary operators describing open systems in the semiclassical limit has a number of new interesting properties and the concept of the fractal Weyl law (Sjöstrand , 1990; Zworski , 1999) has been introduced to describe the dependence of number of resonant Gamow eigenvalues (Gamow , 1928) on $\hbar$.

The Gamow eigenstates find important applications for decay of radioactive nuclei, quantum chemistry reactions, chaotic scattering and microlasers with chaotic resonators, open quantum maps (see (Gaspard, 1998, 2014; Shepelyansky , 2008) and Refs. therein). The spectrum of corresponding operators has a complex spectrum λ. The spread width $\gamma = -2\ln|\lambda|$ of eigenvalues λ determines the life time of a corresponding eigenstate. The understanding of the spectral properties of related operators in the semiclassical limit represents an important challenge.

According to the fractal Weyl law (Lu *et al.*, 2003; Sjöstrand , 1990) the number of Gamow eigenvalues N_γ, which have escape rates γ in a finite band width $0 \le \gamma \le \gamma_b$, scales as

$$N_\gamma \propto \hbar^{-d/2} \propto N^{d/2} \qquad (5)$$

where d is a fractal dimension of a classical strange repeller formed by classical orbits nonescaping in future and past times. In the context of eigenvalues λ of the Google matrix we have $\gamma = -2\ln|\lambda|$. By numerical simulations it has been shown that the law (5) works for a scattering problem in 3-disk system (Lu *et al.*, 2003) and quantum chaos maps with absorption when the fractal dimension d is changed in a broad range $0 < d < 2$ (Ermann and Shepelyansky , 2010b; Shepelyansky , 2008).

The fractal Weyl law (5) of open systems with a fractal dimension $d < 2$ leads to a striking consequence: only a relatively small fraction of eigenvalues $\mu_W \sim N_\gamma/N \propto \hbar^{(2-d)/2} \propto N^{(d-2)/2} \ll 1$ has finite values of $|\lambda|$ while almost all eigenstates of the matrix operator of size $N \propto 1/\hbar$ have $\lambda \to 0$. The eigenstates with finite $|\lambda| > 0$ are related to the classical fractal sets of orbits non-escaping neither in the future neither in the past. A fractal structure of these quantum fractal eigenstates has been investigated in (Shepelyansky , 2008). There it was conjectured that the eigenstates of a Google matrix with finite $|\lambda| > 0$ will select interesting specific communities of a network. We will see below that the fractal Weyl law can indeed be observed in certain directed networks and in particular we show in the next section that it naturally appears for Perron-Frobenius operators of dynamical systems and Ulam networks.

It is interesting to note that nontrivial complex spectra also naturally appear in systems of quantum chaos in presence of a contact with a measurement device (Bruzda *et al.*, 2010). The properties of complex spectra of small

size orthostochastic (unistochastic) matrices are analyzed in (Zyczkowski *et al.*, 2003). In such matrices the elements can be presented in a form $S_{ij} = O_{ij}^2$ ($S_{ij} = |U_{ij}|^2$) where O is an orthogonal matrix (U is a unitary matrix). We will see certain similarities of their spectra with the spectra of directed networks discussed in Sec. VIII.

Recent mathematical results for the fractal Weyl law are presented in (Nonnenmacher and Zworski , 2007; Nonnenmacher *et al.*, 2014).

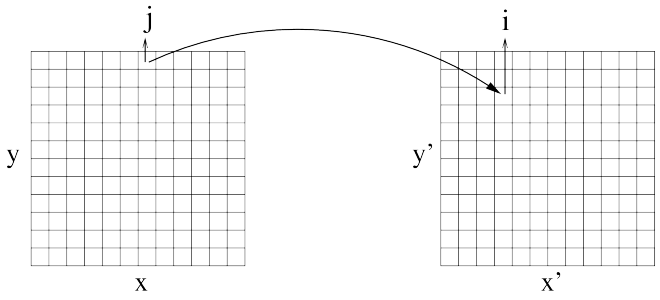

FIG. 8 Illustration of operation of the Ulam method: the phase space (x, y) is divided in $N = N_x \times N_y$ cells, N_c trajectories start from cell j and the number of trajectories N_{ij} arrived to a cell i from a cell j is collected after a map iteration. Then the matrix of Markov transitions is defined as $S_{ij} = N_{ij}/N_c$, by construction $\sum_{i=1}^{N} S_{ij} = 1$.

VI. ULAM NETWORKS

By construction the Google matrix belongs to the class of Perron-Frobenius operators which naturally appear in ergodic theory (Cornfeld*et al.*, 1982) and dynamical systems with Hamiltonian or dissipative dynamics (Brin and Stuck, 2002). In 1960 Ulam (Ulam, 1960) proposed a method, now known as the Ulam method, for a construction of finite size approximants for the Perron-Frobenius operators of dynamical maps. The method is based on discretization of the phase space and construction of a Markov chain based on probability transitions between such discrete cells given by the dynamics. Using as an example a simple chaotic map Ulam made a conjecture that the finite size approximation converges to the continuous limit when the cell size goes to zero. Indeed, it has been proven that for hyperbolic maps in one and higher dimensions the Ulam method converges to the spectrum of continuous system (Blank *et al.*, 2002; Li , 1976). The probability flows in dynamical systems have rich and nontrivial features of general importance, like simple and strange attractors with localized and delocalized dynamics governed by simple dynamical rules (Lichtenberg and Lieberman, 1992). Such objects are generic for nonlinear dissipative dynamics and hence can have relevance for actual WWW structure. The analysis of Ulam networks, generated by the Ulam method, allows to obtain a better intuition about the spectral properties of Google matrix. The term Ulam networks was introduced in (Shepelyansky and Zhirov , 2010a).

A. Ulam method for dynamical maps

In Fig. 8 we show how the Ulam method works. The phase space of a dynamical map is divided in equal cells and a number of trajectories N_c is propagated by a map iteration. Thus a number of trajectories N_{ij} arrived from cell j to cell i is determined. Then the matrix of Markov transition is defined as $S_{ij} = N_{ij}/N_c$. By construction this matrix belongs to the class of Perron-Frobenius operators which includes the Google matrix.

The physical meaning of the coarse grain description by a finite number of cells is that it introduces in the system a noise of cell size amplitude. Due to that an exact time reversibility of dynamical equations of chaotic maps is destroyed due to exponential instability of chaotic dynamics. This time reversibility breaking is illustrated by an example of the Arnold cat map by (Ermann and Shepelyansky , 2012b). For the Arnold cat map on a long torus it is shown that the spectrum of the Ulam approximate of the Perron-Frobenius (UPFO) is composed of a large group of complex eigenvalues with $\gamma \sim 2h \approx 2$, and real eigenvalues with $|1-\lambda| \ll 1$ corresponding to a statistical relaxation to the ergodic state at $\lambda = 1$ described by the Fokker-Planck equation (here h is the Kolmogorov-Sinai entropy of the map being here equal to the Lyapunov exponent, see e.g. (Chirikov , 1979)).

For fully chaotic maps the finite cell size, corresponding to added noise, does not significantly affect the dynamics and the discrete UPFO converges to the limiting case of continuous Perron-Frobenius operator (Blank *et al.*, 2002; Li , 1976). The Ulam method finds useful applications in studies of dynamics of molecular systems and coherent structures in dynamical flows (Froyland and Padberg , 2009). Additional Refs. can be found in (Frahm and Shepelyansky , 2010).

B. Chirikov standard map

However, for symplectic maps with a divided phase space, a noise present in the Ulam method significantly affects the original dynamics leading to a destruction of islands of stable motion and Kolmogorov-Arnold-Moser (KAM) curves. A famous example of such a map is the Chirikov standard map which describes the dynamics of many physical systems (Chirikov , 1979; Chirikov and Shepelyansky , 2008):

$$\bar{y} = \eta y + \frac{K_s}{2\pi} \sin(2\pi x) \ , \ \ \bar{x} = x + \bar{y} \pmod 1 \ . \qquad (6)$$

Here bars mark the variables after one map iteration and we consider the dynamics to be periodic on a torus so that $0 \le x \le 1$, $-1/2 \le y \le 1/2$; K_s is a dimensionless parameter of chaos. At $\eta = 1$ we have area-preserving symplectic map, considered in this SubSec., for $0 < \eta < 1$ we have a dissipative dynamics analyzed in next SubSec.

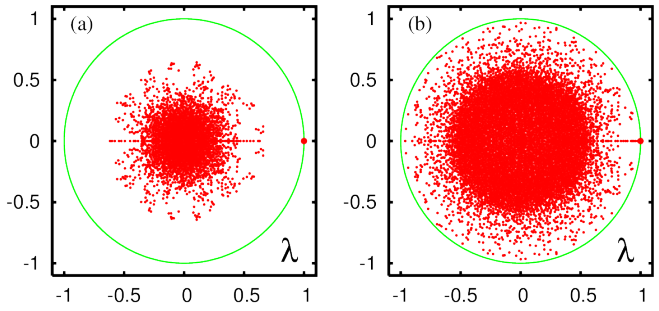

FIG. 9 (Color online) Complex spectrum of eigenvalues λ_j, shown by red/gray dots, for the UPFO of two variants of the Chirikov standard map (6); the unit circle $|\lambda| = 1$ is shown by a green (light gray) curve, the unit eigenvalue at $\lambda = 1$ is shown as larger red/gray dot. Panel (a) corresponds to the Chirikov standard map at dissipation $\eta = 0.3$ and $K_s = 7$; the phase space is covered by 110×110 cells and the UPFO is constructed by many trajectories with random initial conditions generating transitions from one cell into another (after (Ermann and Shepelyansky , 2010b)). Panel (b) corresponds to the Chirikov standard map without dissipation at $K_s = 0.971635406$ with an UPFO constructed from a single trajectory of length 10^{12} in the chaotic domain and $280 \times 280/2$ cells to cover the phase space (after (Frahm and Shepelyansky , 2010)).

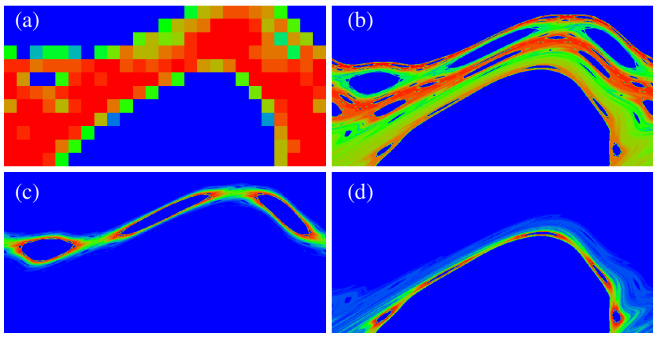

FIG. 10 (Color online) Density plots of absolute values of the eigenvectors of the UPFO obtained by the generalized Ulam method with a single trajectory of 10^{12} iterations of the Chirikov standard map at $K_s = 0.971635406$. The phase space is shown in the area $0 \le x \le 1$, $0 \le y \le 1/2$; the UPFO is obtained from $M \times M/2$ cells placed in this area. Panels represent: (a) eigenvector ψ_0 with eigenvalue $\lambda_0 = 1$; (b) eigenvector ψ_2 with real eigenvalue $\lambda_2 = 0.99878108$; (c) eigenvector ψ_6 with complex eigenvalue $\lambda_6 = -0.49699831 + i\,0.86089756 \approx |\lambda_6|\, e^{i\,2\pi/3}$; (d) eigenvector ψ_{13} with complex eigenvalue $\lambda_{13} = 0.30580631 + i\,0.94120900 \approx |\lambda_{13}|\, e^{i\,2\pi/5}$. Panel (a) corresponds to $M = 25$ while (b), (c) and (d) have $M = 800$. Color is proportional to amplitude with blue (black) for zero and red (gray) for maximal value. After (Frahm and Shepelyansky , 2010).

Since the finite cell size generates noise and destroys the KAM curves in the map (6) at $\eta = 1$, one should use the generalized Ulam method (Frahm and Shepelyansky , 2010), where the transition probabilities N_{ij}/N_c are collected along one chaotic trajectory. In this construction a trajectory visits only those cells which belong to one connected chaotic component. Therefore the noise induced by the discretization of the phase space does not lead to a destruction of invariant curves, in contrast to the original Ulam method (Ulam, 1960), which uses all cells in the available phase space. Since a trajectory is generated by a continuous map it cannot penetrate inside the stability islands and on a physical level of rigor one can expect that, due to ergodicity of dynamics on one connected chaotic component, the UPFO constructed in such a way should converge to the Perron-Frobenius operator of the continuous map on a given subspace of chaotic component. The numerical confirmations of this convergence are presented in (Frahm and Shepelyansky , 2010).

We consider the map (6) at $K_s = 0.971635406$ when the golden KAM curve is critical. Due to the symmetry of the map with respect to $x \to 1-x$ and $y \to -y$ we can use only the upper part of the phase space with $y \ge 0$ dividing it in $M \times M/2$ cells. At that K_s we find that the number of cells visited by the trajectory in this half square scales as $N_d \approx C_d M^2/2$ with $C_d \approx 0.42$. This means that the chaotic component contains about 40% of the total area which is in good agreement with the known result of (Chirikov , 1979).

The spectrum of the UPFO matrix S for the phase space division by $280 \times 208/2$ cells is shown in Fig. 9(b). In a first approximation the spectrum λ of S is more or less homogeneously distributed in the polar angle φ defined as $\lambda_j = |\lambda_j| \exp(i\varphi_j)$. With the increase of matrix size N_d the two-dimensional density of states $\rho(\lambda)$ converges to a limiting distribution (Frahm and Shepelyansky , 2010). With the help of the Arnoldi method it is possible to compute a few thousands of eigenvalues with largest absolute values $|\lambda|$ for maximal $M = 1600$ with the total matrix size $N = N_d \approx 5.3 \times 10^5$.

The eigenstate at $\lambda = 1$ is homogeneously distributed over the chaotic component at $M = 25$ (Fig. 10) and higher M values (Frahm and Shepelyansky , 2010). This results from the ergodicity of motion and the fact that for symplectic maps the measure is proportional to the phase space area (Chirikov , 1979; Cornfeld *et al.*, 1982). Examples of other right eigenvalues of S at real and complex eigenvalues λ with $|\lambda| < 1$ are also shown in Fig. 10. For λ_2 the eigenstate corresponds to some diffusive mode with two nodal lines, while other two eigenstates are localized around certain resonant structures in phase space. This shows that eigenstates of the matrix G (and S) are related to specific communities of a network.

With the increase of number of cells $M^2/2$ there are eigenvalues which become more and more close to the unit eigenvalue. This is shown to be related to an algebraic statistics of Poincaré recurrences and long time sticking of trajectories in a vicinity of critical KAM curves. At the same time for symplectic maps the measure is proportional to area so that we have dimension $d = 2$ and hence we have a usual Weyl law with $N_\gamma \propto N$.

More details can be found at (Frahm and Shepelyansky , 2010, 2013).

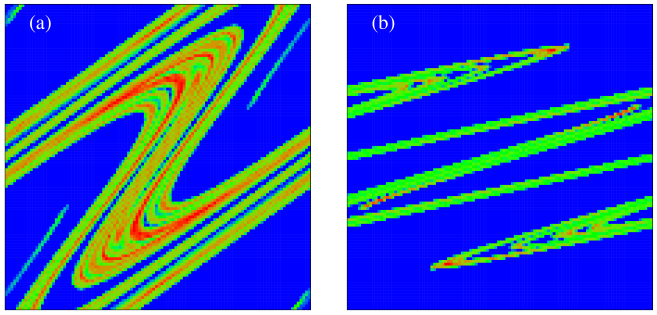

FIG. 11 (Color online) Phase space representation of eigenstates of the UFPO S for $N = 110 \times 110$ cells (color is proportional to absolute value $|\psi_i|$ with red/gray for maximum and blue/black for zero). Panel (a) shows an eigenstate with maximum eigenvalue $\lambda_1 = 0.756$ of the UFPO of map (6) with absorption at $K_s = 7$, $a = 2$, $\eta = 1$, the space region is $(-aK_s/4\pi \le y \le aK_s/4\pi,\ 0 \le x \le 1)$, the fractal dimension of the strange repeller set nonescaping in future is $d_e = 1 + d/2 = 1.769$. Panel (b) shows an eigenstate at $\lambda = 1$ of the UFPO of map (6) without absorption at $K_s = 7$, $\eta = 0.3$, the shown space region is $(-1/\pi \le y \le 1/\pi,\ 0 \le x \le 1)$ and the fractal dimension of the strange attractor is $d = 1.532$. After (Ermann and Shepelyansky , 2010b).

C. Dynamical maps with strange attractors

The fractal Weyl law (5) has initially been proposed for quantum systems with chaotic scattering. However, it is natural to assume that it should also work for Perron-Frobenius operators of dynamical systems. Indeed, the mathematical results for the Selberg zeta function indicated that the law (5) should remain valid for the UFPO (see Refs. at (Nonnenmacher *et al.*, 2014)). A detailed test of this conjecture (Ermann and Shepelyansky , 2010b) has been performed for the map (6) with dissipation at $0 < \eta < 1$, when at large K_s the dynamics converges to a strange attractor in the range $-2 < y < 2$, and for the nondissipative case $\eta = 1$ with absorption where all orbits leaving the interval $-aK_s/4\pi \le y \le aK_s/4\pi$ are absorbed after one iteration (in both cases there is no modulus in y).

An example of the spectrum of UPFO for the model with dissipation is shown in Fig. 9(a). We see that now, in contrast to the symplectic case of Fig. 9(b), the spectrum has a significant gap which separates the eigenvalue $\lambda = 1$ from the other eigenvalues with $|\lambda| < 0.7$. For the case with absorption the spectrum has a similar structure but now with $|\lambda| < 1$ for the leading eigenvalue λ since the total number of initial trajectories decreases with the number of map iterations due to absorption implying that for this case $\sum_i S_{ij} < 1$ with S being the UPFO.

It is established that the distribution of density of states $dW/d\gamma$ (or $dW/d|\lambda|$) converges to a fixed distribution in the limit of large N or cell size going to zero (Ermann and Shepelyansky , 2010b) (see Fig.4 there). This demonstrates the validity of the Ulam conjecture for considered systems.

Examples of two eigenstates of the UFPO for these two models are shown in Fig. 11. The fractal structure of eigenstates is well visible. For the dissipative case without absorption we have eigenstates localized on the strange attractor. For the case with absorption eigenstates are located on a strange repeller corresponding to an invariant set of nonescaping orbits. The fractal dimension d of these classical invariant sets can be computed by the usual box-counting method for dynamical systems. It is important to note that for the case with absorption it is more natural to measure the dimension d_e of the set of orbits nonescaping in future. Due to the time reversal symmetry of the continuous map the dimension of the set of orbits nonescaping in the past is also d_e. Thus the phase space dimension 2 is composed of $2 = d_e + d_e - d$ and $d_e = 1 + d/2$ where d is the dimension of the invariant set of orbits nonescaping neither in the future neither in the past. For the case with dissipation without absorption all orbits drop on a strange attractor and we have the dimension of invariant set $d_e = d$.

D. Fractal Weyl law for Perron-Frobenius operators

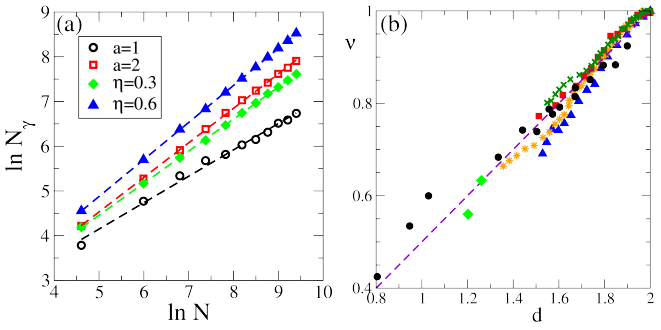

FIG. 12 (Color online) Panel (a) shows the dependence of the integrated number of states N_γ with decay rates $0 \le \gamma \le \gamma_b = 16$ on the size N of the UFPO matrix S for the map (6) at $K_s = 7$. The fits of numerical data, shown by dashed straight lines, give $\nu = 0.590, d_e = 1 + d/2 = 1.643$ (at $a = 1$); $\nu = 0.772, d_e = 1 + d/2 = 1.769$ (at $a = 2$); $\nu = 0.716, d = 1.532$ (at $\eta = 0.3$); $\nu = 0.827, d = 1.723$ (at $\eta = 0.6$). Panel (b) shows the fractal Weyl exponent ν as a function of fractal dimension d of the invariant fractal set for the map (6) with a strange attractor ($\eta < 1$) at $K_s = 15$ (green/gray crosses), $K_s = 12$ (red/gray squares), $K_s = 10$ (orange/gray stars), $K_s = 7$ blue/black triangles; for a strange repeller ($\eta = 1$) at $K_s = 7$ (black points) and for a strange attractor for the Hénon map at standard parameters $a = 1.2; 1.4$, $b = 0.3$ (green diamonds). The straight dashed line shows the fractal Weyl law dependence $\nu = d/2$. After (Ermann and Shepelyansky , 2010b).

The direct verification of the validity of the fractal Weyl law (5) is presented in Fig. 12. The number of eigenvalues N_γ in a range with $0 \le \gamma \le \gamma_b$ ($\gamma = -2\ln|\lambda|$) is numerically computed as a function of matrix size N. The fit of the dependence $N_\gamma(N)$, as shown in Fig. 12(a), allows to determine the exponent ν in the relation $N_\gamma \propto N^\nu$. The dependence of ν on the fractal dimension d, computed from the invariant fractal set by the box-counted method, is shown in Fig. 12(b). The numerical data are in good agreement with the theoretical fractal Weyl law dependence $\nu = d/2$. This law works for a variety of parameters for the system (6) with absorption and dissipation, and also for a strange attractor in the Hénon map ($\bar{x} = y + 1 - ax^2, \bar{y} = bx$). We attribute certain deviations, visible in Fig. 12 especially for $K_s = 7$, to the fact that at $K_s = 7$ there is a small island of stability at $\eta = 1$, which can produce certain influence on the dynamics.

The physical origin of the law (5) can be understood in a simple way: the number of states N_γ with finite values of γ is proportional to the number of cells $N_f \propto N^{d/2}$ on the fractal set of strange attractor. Indeed, the results for the overlap measure show that the eigenstates N_γ have a strong overlap with the steady state while the states with $\lambda \to 0$ have very small overlap. Thus almost all N states have eigenvalues $\lambda \to 0$ and only a small fraction of states on a strange attractor/repeller $N_\gamma \propto N_f \propto N^{d/2} \ll N$ has finite values of λ. We also checked that the participation ratio ξ of the eigenstate at $\lambda = 1$, grows as $\xi \sim N_f \propto N^{d/2}$ in agreement with the fractal Weyl law (Ermann and Shepelyansky , 2010b).

E. Intermittency maps

The properties of the Google matrix generated by one-dimensional intermittency maps are analyzed in (Ermann and Shepelyansky , 2010a). It is found that for such Ulam networks there are many eigenstates with eigenvalues $|\lambda|$ being very close to unity. The PageRank of such networks at $\alpha = 1$ is characterized by a power law decay with an exponent determined by the parameters of the map. It is interesting to note that usually for WWW the PageRank probability is proportional to a number of ingoing links distribution (see e.g. (Litvak *et al.*, 2008)). For the case of intermittency maps the decay of PageRank is independent of number of ingoing links. In addition, for α close to unity a decay of the PageRank has an exponent $\beta \approx 1$ but at smaller values $\alpha \le 0.9$ the PageRank becomes completely delocalized. It is shown that the delocalization depends on the intermittency exponent of the map. This indicates that a rather dangerous phenomenon of PageRank delocalization can appear for certain directed networks. At the same time the one-dimensional intermittency map still generates a relatively simple structure of links with a typical number of links per node being close to unity. Such a case is probably not very typical for real networks. Therefore it is useful to analyze richer Ulam networks with a larger number of links per node.

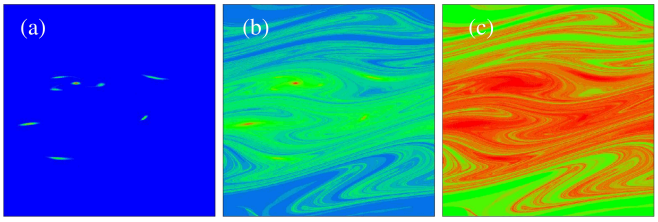

FIG. 13 (Color online) PageRank probability P_j for the Google matrix generated by the Chirikov typical map at $T = 10$, $k_s = 0.22$, $\eta = 0.99$ with $\alpha = 1$ (a), $\alpha = 0.95$ (b), and $\alpha = 0.85$ (c). The probability P_j is shown in the phase space region $0 \le x < 2\pi; -\pi \le y < \pi$ which is divided in $N = 3.6 \cdot 10^5$ cells; P_j is zero for blue/black and maximal for red/gray. After (Shepelyansky and Zhirov , 2010a).

F. Chirikov typical map

With this aim we consider the Ulam networks generated by the Chirikov typical map with dissipation studied by (Shepelyansky and Zhirov , 2010a). The map introduced, by Chirikov in 1969 for description of continuous chaotic flows, has the form:

$$y_{t+1} = \eta y_t + k_s \sin(x_t + \theta_t) \ , \quad x_{t+1} = x_t + y_{t+1} \ . \qquad (7)$$

Here the dynamical variables x, y are taken at integer moments of time t. Also x has a meaning of phase variable and y is a conjugated momentum or action. The phases $\theta_t = \theta_{t+T}$ are T random phases periodically repeated along time t. We stress that their T values are chosen and fixed once and they are not changed during the dynamical evolution of x, y. We consider the map in the region of Fig. 13 ($0 \le x < 2\pi, -\pi \le y < \pi$) with the 2π-periodic boundary conditions. The parameter $0 < \eta < 1$ gives a global dissipation. The properties of the symplectic map at $\eta = 1$ have been studied in detail in (Frahm and Shepelyansky , 2009). The dynamics is globally chaotic for $k_s > k_c \approx 2.5/T^{3/2}$ and the Kolmogorov-Sinai entropy is $h \approx 0.29 {k_s}^{2/3}$ (more details about the Kolmogorov-Sinai entropy can be found in (Brin and Stuck, 2002; Chirikov , 1979; Cornfeld*et al.*, 1982)). A bifurcation diagram at $\eta < 1$ shows a series of transitions between fixed points, simple and strange attractors. Here we present results for $T = 10$, $k_s = 0.22$, $\eta = 0.99$ and a specific random set of θ_t given in (Shepelyansky and Zhirov , 2010a).

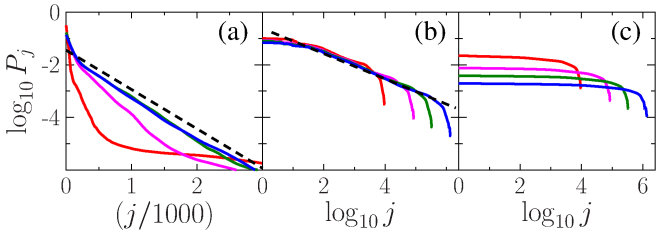

FIG. 14 (Color online) Dependence of PageRank probability P_j on PageRank index j for number of cells in the UFPO being $N = 10^4$, 9×10^4, 3.6×10^5 and 1.44×10^6 (larger N have more dark and more long curves in (b), (c); in (a) this order of N is for curves from bottom to top (curves for $N = 3.6 \times 10^5$ and 1.44×10^6 practically coincide in this panel; for online version we note that the above order of N values corresponds to red, magenta, green, blue curves respectively). Dashed line in (a) shows an exponential Boltzmann decay (see text, line is shifted in j for clarity). The dashed straight line in (b) shows the fit $P_j \sim 1/j^\beta$ with $\beta = 0.48$. Other parameters, including the values of α, and panel order are as in Fig. 13. After (Shepelyansky and Zhirov , 2010a).

Due to exponential instability of motion one cell in the Ulam method gives transitions approximately to $k_{\rm cl} \approx \exp(hT)$ other cells. According to this relation a large number of cells $k_{\rm cl}$ can be coupled at large T and h. For parameters of Fig. 13 one finds an approximate power law distribution of ingoing and outgoing links in the corresponding Ulam network with the exponents $\mu_{\rm in} \approx \mu_{\rm out} \approx 1.9$. The variation of the PageRank vector with the damping factor α is shown in Fig. 13 on the phase plane (x, y). For $\alpha = 1$ the PageRank is concentrated in a vicinity of a simple attractor composed of several fixed points on the phase plane. Thus the dynamical attractors are the most popular nodes from the network view point. With a decrease of α down to $0.95, 0.85$ values we find a stronger and stronger delocalization of PageRank over the whole phase space.

The delocalization with a decrease of α is also well seen in Fig. 14 where we show P_j dependence on PageRank index j with a monotonic decreasing probability P_j. At $\alpha = 1$ we have an exponential decay of P_j with j that corresponds to a Boltzmann type distribution where a noise produced by a finite cell size in the Ulam method is compensated by dissipation. For $\alpha = 0.95$ the random jumps of a network surfer, induced by the term $(1 - \alpha)/N$ in (1), produce an approximate power law decay of $P_j \propto 1/j^\beta$ with $\beta \approx 0.48$. For $\alpha = 0.85$ the PageRank probability is flat and completely delocalized over the whole phase space.

The analysis of the spectrum of S for the map (7) for the parameters of Fig. 14 shows the existence of eigenvalues being very close to $\lambda = 1$, however, there is no exact degeneracy as it is the case for UK universities which we will discuss below. The spectrum is characterized by the fractal Weyl law with the exponent $\nu \approx 0.85$. For eigenstates with $|\lambda| < 1$ the values of IPR ξ are less than 300 for a matrix size $N \approx 1.4 \times 10^4$ showing that eigenstates are localized. However, for the PageRank the computations can be done with larger matrix sizes reaching a maximal value of $N = 6.4 \times 10^5$. The dependence of ξ on α shows that a delocalization transition of PageRank vector takes place for $\alpha < \alpha_c \approx 0.95$. Indeed, at $\alpha = 0.98$ we have $\xi \approx 30$ while at $\alpha \approx 0.8$ the IPR value of PageRank becomes comparable with the whole system size $\xi \approx 5 \times 10^5 \sim N = 6.4 \times 10^5$ (see Fig.9 at (Shepelyansky and Zhirov , 2010a)).

The example of Ulam networks considered here shows that a dangerous phenomenon of PageRank delocalization can take place under certain conditions. This delocalization may represent a serious danger for efficiency of search engines since for a delocalized flat PageRank the ranking of nodes becomes very sensitive to small perturbations and fluctuations.

VII. LINUX KERNEL NETWORKS

Modern software codes represent now complex large scale structures and analysis and optimization of their architecture become a challenge. An interesting approach to this problem, based on a directed network construction, has been proposed by (Chepelianskii, 2010). Here we present results obtained for such networks.

A. Ranking of software architecture

Following (Chepelianskii, 2010) we consider the Procedure Call Networks (PCN) for open source programs with emphasis on the code of Linux Kernel (Linux, 2010) written in the C programming language (Kernighan and Ritchie, 1978). In this language the code is structured as a sequence of procedures calling each other. Due to that feature the organization of a code can be naturally represented as a PCN, where each node represents a procedure and each directed link corresponds to a procedure call. For the Linux source code such a directed network is built by its lexical scanning with the identification of all the defined procedures. For each of them a list keeps track of the procedures calls inside their definition.

An example of the obtained network for a toy code with two procedures *start_kernel* and *printk* is shown in Fig. 15. The in/out-degrees of this model, noted as k and $\bar{k}$, are shown in Fig. 15. These numbers correspond to the number of out/in-going calls for each procedure. The obtained in/out-degree probability distributions $P_{\rm in}(k)$, $P_{\rm out}(\bar{k})$ are shown Fig. 15 for different Linux Kernel releases. These distributions are well described by power law dependencies $P_{\rm in}(k) \propto 1/k^{\mu_{\rm in}}$ and $P_{\rm out}(\bar{k}) \propto 1/\bar{k}^{\mu_{\rm out}}$ with $\mu_{\rm in} = 2.0 \pm 0.02$, and $\mu_{\rm out} = 3.0 \pm 0.1$. These values of exponents are close to those found for the WWW (Donato *et al.*, 2004; Pandurangan *et al.*, 2005). If only calls to distinct functions are counted in the outdegree distribution then the exponent drops to $\mu_{\rm out} \approx 5$ whereas $\mu_{\rm in}$ remains unchanged. It is important that the distribu-

tions for the different kernel releases remain stable even if the network size increases from $N = 2751$ for version V1.0 to $N = 285509$ for the latest version V2.6.32 taken into account in this study. This confirms the free-scale structure of software architecture of Linux Kernel network.

The probability distributions of PageRank and CheiRank vectors are also well described by power laws with exponents $\beta_{\rm in} \approx 1$ and $\beta_{\rm out} \approx 0.5$ being in good agreement with the usual relation $\beta = 1/(\mu - 1)$ (see Fig.2 in (Chepelianskii, 2010)). For V2.6.32 the top three procedures of PageRank at $\alpha = 0.85$ are *printk, memset, kfree* with probabilities $0.024, 0.012, 0.011$ respectively, while at the top of CheiRank we have *start_kernel, btrfs_ioctl, menu_finalize* with respectively $0.000280, 0.000255, 0.000250$. These procedures perform rather different tasks with *printk* reporting messages and *start_kernel* initializing the Kernel and managing the repartition of tasks. This gives an idea that both PageRank and CheiRank order can be useful to highlight en different aspects of directed and inverted flows on our network. Of course, in the context of WWW ingoing links related to PageRank are less vulnerable as compared to outgoing links related to CheiRank, which can be modified by a user rather easily. However, in other type of networks both directions of links appear in a natural manner and thus both vectors of PageRank and CheiRank play an important and useful role.

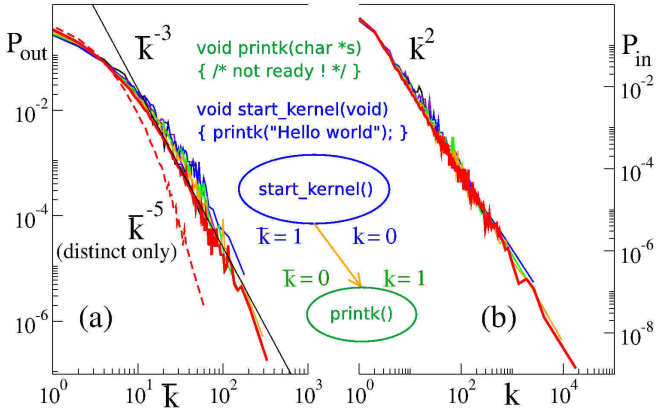

FIG. 15 (Color online) The diagram in the center represents the PCN of a toy kernel with two procedures written in C-programming language. The data on panels (a) and (b) show outdegree and indegree probability distributions $P_{\rm out}(\bar{k})$ and $P_{\rm in}(k)$ respectively. The colors correspond to different Kernel releases. The most recent version 2.6.32, with $N = 285509$ and an average 3.18 calls per procedure, is represented in red/gray. Older versions (2.4.37.6, 2.2.26, 2.0.40, 1.2.12, 1.0) with N respectively equal to (85756, 38766, 14079, 4358, 2751) follow the same behavior. The dashed curve in (a) shows the outdegree probability distribution if only calls to distinct destination procedures are kept. After (Chepelianskii, 2010).

For the Linux Kernel network the correlator κ (4) between PageRank and CheiRank vectors is close to zero (see Fig. 6). This confirms the independence of two vectors. The density distribution of nodes of the Linux Kernel network, shown in Fig. 7(b), has a homogeneous distribution along $\ln K + \ln K^* = const$ lines demonstrating once more absence of correlations between $P(K_i)$ and $P^*(K_i{}^*)$. Indeed, such homogeneous distributions appear if nodes are generated randomly with factorized probabilities $P_i P_i{}^*$ (Chepelianskii, 2010; Zhirov *et al.*, 2010). Such a situation seems to be rather generic for software architecture. Indeed, other open software codes also have a small values of correlator, e.g. OpenSource software including Gimp 2.6.8 has $\kappa = -0.068$ at $N = 17540$ and X Windows server R7.1-1.1.0 has $\kappa = -0.027$ at $N = 14887$. In contrast to these software codes the Wikipedia networks have large values of κ and inhomogeneous distributions in (K, K^*) plane (see Figs. 6,7).

The physical reasons for absence of correlations between $P(K)$ and $P^*(K^*)$ have been explained in (Chepelianskii, 2010) on the basis of the concept of "separation of concerns" in software architecture (Dijkstra, 1982). It is argued that a good code should decrease the number of procedures that have high values of both PageRank and CheiRank since such procedures will play a critical role in error propagation since they are both popular and highly communicative at the same time. For example in the Linux Kernel, *do_fork*, that creates new processes, belongs to this class. Such critical procedures may introduce subtle errors because they entangle otherwise independent segments of code. The above observations suggest that the independence between popular procedures, which have high $P(K_i)$ and fulfill important but well defined tasks, and communicative procedures, which have high $P^*(K_i{}^*)$ and organize and assign tasks in the code, is an important ingredient of well structured software.

B. Fractal dimension of Linux Kernel Networks

The spectral properties the Linux Kernel network are analyzed in (Ermann *et al.*, 2011a). At large N the spectrum is obtained with the help of Arnoldi method from ARPACK library. This allows to find eigenvalues with $|\lambda| > 0.1$ for the maximal N at V2.6.32. An example of complex spectrum λ of G is shown in Fig. 16(a). There are clearly visible lines at real axis and polar angles $\varphi = \pi/2, 2\pi/3, 4\pi/3, 3\pi/2$. The later are related to certain cycles in procedure calls, e.g. an eigenstate at $\lambda_i = 0.85\exp(i2\pi/3)$ is located only on 6 nodes. The spectrum of G^* has a similar structure.

The network size N grows with the version number of Linux Kernel corresponding to its evolution in time. We determine the total number of states N_λ with $0.1 < |\lambda| \leq 1$ and $0.25 < |\lambda| \leq 1$. The dependence of N_λ on N, shown in Fig. 16(b), clearly demonstrates the validity of the fractal Weyl law with the exponent $\nu \approx 0.63$ for G (we find $\nu^* \approx 0.65$ for G^*). We take the values of ν for $\lambda = 0.1$ where the number of eigenvalues N_λ gives a

better statistics. Within statistical errors the value of ν is not sensitive to the cutoff value at small λ. The matrix G^* has slightly higher values of ν. These results show that the PCN of Linux Kernel has a fractal dimension $d = 2\nu \approx 1.26$ for G and $d = 2\nu \approx 1.3$ for G^*.

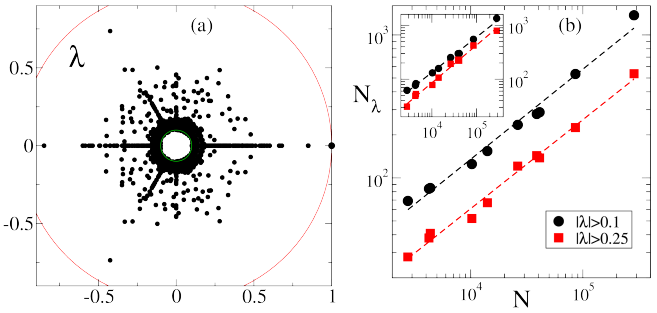

FIG. 16 (Color online) Panel (a) shows distribution of eigenvalues λ in the complex plane for the Google matrix G of the Linux Kernel version 2.6.32 with $N = 285509$ and $\alpha = 0.85$; the solid curves represent the unit circle and the lowest limit of computed eigenvalues. Panel (b) shows dependence of the integrated number of eigenvalues N_λ with $|\lambda| > 0.25$ (red/gray squares) and $|\lambda| > 0.1$ (black circles) as a function of the total number of processes N for versions of Linux Kernels. The values of N correspond (in increasing order) to Linux Kernel versions 1.0, 1.1, 1.2, 1.3, 2.0, 2.1, 2.2, 2.3, 2.4 and 2.6. The power law $N_\lambda \propto N^\nu$ has fitted values $\nu_{|\lambda|>0.25} = 0.622 \pm 0.010$ and $\nu_{|\lambda|>0.1} = 0.630 \pm 0.015$. Inset shows data for the Google matrix G^* with inverse link directions, the corresponding exponents are $\nu^*_{|\lambda|>0.25} = 0.696 \pm 0.010$ and $\nu^*_{|\lambda|>0.1} = 0.652 \pm 0.007$. After (Ermann *et al.*, 2011a).

To check that the fractal dimension of the PCN indeed has this value the dimension of the network is computed by another direct method known as the cluster growing method (see e.g. (Song *et al.*, 2005)). In this method the average mass or number of nodes $\langle M_c \rangle$ is computed as a function of the *network distance* l counted from an initial seed node with further averaging over all seed nodes. For a dimension d the mass $\langle M_c \rangle$ should grow as $\langle M_c \rangle \propto l^d$ that allows to determine the value of d for a given network. It should be noted that the above method should be generalized for the case of directed networks. For that the network distance l is computed following only outgoing links. The average of $\langle M_c(l) \rangle$ is done over all nodes. Due to global averaging the method gives the same result for the matrix with inverted link direction (indeed, the total number of outgoing links is equal to the number of ingoing links). However, as established in (Ermann *et al.*, 2011a), the fractal dimension obtained by this generalized method is very different from the case of converted undirected network, when each directed link is replaced by an undirected one. The average dimension obtained with this method for PCN is $d = 1.4$ even if a certain 20% increase of d appears for the latest Linux versions V2.6. We attribute this deviation for the version V2.6 to the well known fact that significant rearrangements in the Linux Kernel have been done after version V2.4 (Linux, 2010).

Thus in view of the above restrictions we consider that there is a rather good agreement of the fractal dimension obtained from the fractal Weyl law with $d \approx 1.3$ and the value obtained with the cluster growing method which gives an average $d \approx 1.4$. The fact that d is approximately the same for all versions up to V2.4 means that the Linux Kernel is characterized by a self-similar fractal growth in time. The closeness of d to unity signifies that procedure calls are almost linearly ordered that corresponds to a good code organization. Of course, the fractal Weyl law gives the dimension d obtained during time evolution of the network. This dimension is not necessary the same as for a given version of the network of fixed size. However, one can expect that the growth goes in a self-similar way (Dorogovtsev *et al.*, 2008) and that the static dimension is close to the dimension value emerging during the time evolution. This can be viewed as a some kind of ergodicity conjecture. Our data show that this conjecture works with a good accuracy up to the Linux Kernel V.2.6.

Thus the results obtained in (Ermann *et al.*, 2011a) and described here confirm the validity of the fractal Weyl law for the Linux Kernel network with the exponent $\nu \approx 0.65$ and the fractal dimension $d \approx 1.3$. It is important to note that the fractal Weyl exponent ν is not sensitive to the exponent β characterizing the decay of the PageRank. Indeed, the exponent β remains practically the same for the WWW (Donato *et al.*, 2004) and the PCN of Linux Kernel (Chepelianskii, 2010) while the values of fractal dimension are different with $d \approx 4$ for WWW and $d \approx 1.3$ for PCN (see (Ermann *et al.*, 2011a) and Refs. therein).

The analysis of the eigenstates of G and G^* shows that their IPR values remain small ($\xi < 70$) compared to the matrix size $N \approx 2.8 \times 10^5$ showing that they are well localized on certain selected nodes.

VIII. WWW NETWORKS OF UK UNIVERSITIES

The WWW networks of certain UK universities for years between 2002 and 2006 are publicly available at (UK universities, 2011). Due to their modest size, these networks are well suitable for a detail study of PageRank, CheiRank, complex eigenvalue spectra and eigenvectors (Frahm *et al.*, 2011).

A. Cambridge and Oxford University networks

We start our analysis of WWW university networks from those of Cambridge and Oxford 2006. For example, in Fig. 5 we show the dependence of PageRank (CheiRank) probabilities $P(P^*)$ on rank index K (K^*) for the WWW of Cambridge 2006 at $\alpha = 0.85$. The decay is satisfactory described by a power law with the exponent $\beta = 0.75$ ($\beta = 0.61$).

The complex eigenvalue spectrum and the invariant subspace structure (see section III.C) have been studied in great detail for the cases of Cambridge 2006 and Oxford 2006. For Cambridge 2006 (Oxford 2006) the network size is $N = 212710$ (200823) and the number of links is $N_\ell = 2015265$ (1831542). There are $n_{\rm inv} = 1543(1889)$ invariant subspaces, with maximal dimension $d_{\rm max} = 4656(1545)$, together they contain $N_s = 48239$ (30579) subspace nodes leading to 3508 (3275) eigenvalues (of the matrix S) with $|\lambda_j| = 1$ of which $n_1 = 1832(2360)$ are at $\lambda_j = 1$ (about 1% of N). The last number n_1 is larger than the number of invariant subspaces $n_{\rm inv}$ since each of the subspaces has at least one unit eigenvalue because each subspace is described by a full representation matrix of the Perron-Frobenius type. To determine the complex eigenvalue spectrum one can apply exact diagonalization on each subspace and the Arnoldi method on the remaining core space.

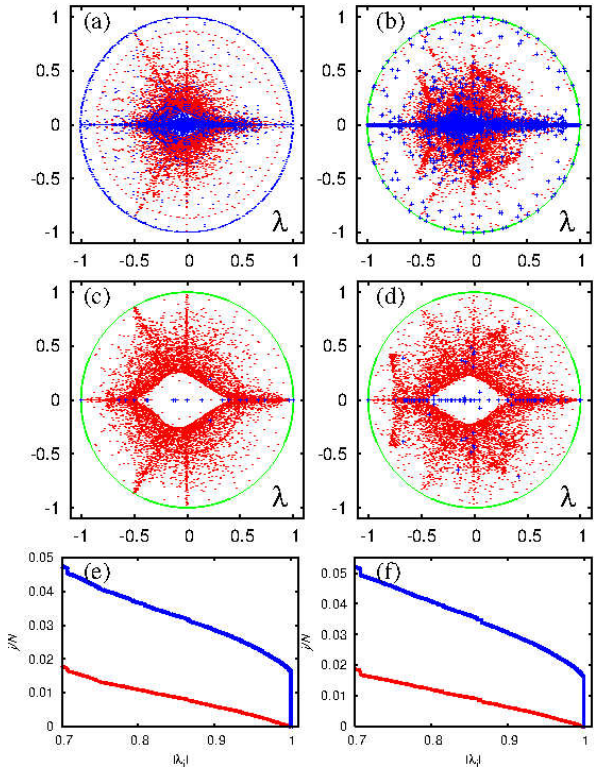

FIG. 17 (Color online) Panels (a) and (b) show the complex eigenvalue spectrum λ of matrix S for the University of Cambridge 2006 and Oxford 2006 respectively. The spectrum λ of matrix S^* for Cambridge 2006 and Oxford 2006 are shown in panels (c) and (d). Eigenvalues λ of the core space are shown by red/gray points, eigenvalues of isolated subspaces are shown by blue/black points and the green/gray curve (when shown) is the unit circle. Panels (e) and (f) show the fraction j/N of eigenvalues with $|\lambda| > |\lambda_j|$ for the core space eigenvalues (red/gray bottom curve) and all eigenvalues (blue/black top curve) from top row data for Cambridge 2006 and Oxford 2006. After (Frahm *et al.*, 2011).

The spectra of all subspace eigenvalues and $n_A = 20000$ core space eigenvalues of the matrices S and S^* are shown in Fig. 17. Even if the decay of PageRank and CheiRank probabilities with rank index is rather similar for both universities (see Fig.1 in (Frahm *et al.*, 2011)) the spectra of two networks are very different. Thus the spectrum contains much more detailed information about the network features compared to the rank vectors.

At the same time the spectra of two universities have certain similar features. Indeed, one can identify cross and triple-star structures. These structures are very similar to those seen in the spectra of random orthostochastic matrices of small size $N = 3, 4$ shown in Fig. 18 from (Zyczkowski *et al.*, 2003) (spectra of unistochastic matrices have a similar structure). The spectrum borders, determined analytically in (Zyczkowski *et al.*, 2003) for these N values, are also shown. The similarity is more visible for the spectrum of S^* case ((c) and (d) of Fig. 17). We attribute this to a larger randomness in outgoing links which have more fluctuations compared to ingoing links, as discussed in (Eom *et al.* , 2013b). The similarity of spectra of Fig. 17 with those of random matrices in Fig. 18 indicates that there are dominant triple and quadruple structures of nodes present in the University networks which are relatively weakly connected to other nodes.

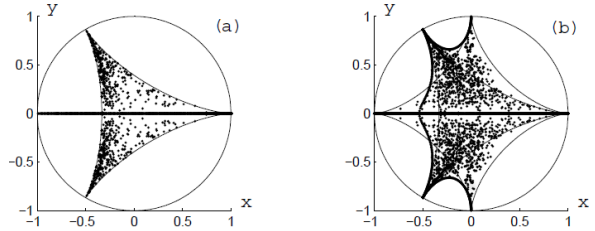

FIG. 18 Spectra λ of 800 random orthostochastic matrices of size $N = 3$ (a) and $N = 4$ (b) ($Re\lambda = x, Im\lambda = y$). Thin lines denote 3- and 4-hypocycloids, while the thick lines represent the 3-4 interpolation arc. After (Zyczkowski *et al.*, 2003).

The core space submatrix S_{cc} of Eq. (2) does not obey to the column sum normalization due to non-vanishing elements in the block S_{sc} which allow for a small but finite escape probability from core space to subspace nodes. Therefore the maximum eigenvalue of the core space (of the matrix S_{cc}) is below unity. For Cambridge 2006 (Oxford 2006) it is given by $\lambda_1^{(\rm core)} = 0.999874353718$ (0.999982435081) with a quite clear gap $1-\lambda_1^{(\rm core)} \sim 10^{-4}$ ($\sim 10^{-5}$).

B. Universal emergence of PageRank

For $\alpha = 1$ the leading eigenvalue $\lambda = 1$ is highly degenerate due to the subspace structure. This degeneracy is lifted for $\alpha < 1$ with a unique eigenvector, the PageRank, for the leading eigenvalue. The question arises how the PageRank emerges if $1 - \alpha \ll 1$. Following (Frahm

et al., 2011), an answer is obtained from a formal matrix expression:

$$P = (1-\alpha)\,(I-\alpha S)^{-1}\,e/N, \qquad (8)$$

where the vector e has unit entries on each node and I is the unit matrix. Then, assuming that S is diagonalizable (with no nontrivial Jordan blocks) we can use the expansion:

$$P = \sum_{\lambda_j=1} c_j\,\psi_j + \sum_{\lambda_j\neq 1} \frac{1-\alpha}{(1-\alpha)+\alpha(1-\lambda_j)}\,c_j\,\psi_j\ . \qquad (9)$$

where ψ_j are the eigenvectors of S and c_j coefficients determined by the expansion $e/N = \sum_j c_j\psi_j$. Thus Eq. (9) indicates that in the limit $\alpha \to 1$ the PageRank converges to a particular linear combination of the eigenvectors with $\lambda_j = 1$, which are all localized in one of the subspaces. For a finite but very small value of $1-\alpha \ll 1-\lambda_1^{(\rm core)}$ the corrections for the contributions of the core space nodes are $\sim (1-\alpha)/(1-\lambda_1^{(\rm core)})$. This behavior is indeed confirmed by Fig. 19 (a) showing the evolution of the PageRank for different values of $1-\alpha$ for the case of Cambridge 2006 and using a particular method, based on an alternate combination of the power iteration method and the Arnoldi method (Frahm *et al.*, 2011), to determine numerically the PageRank for very small values of $1-\alpha \sim 10^{-8}$.

However, for certain of the university networks the core space gap $1-\lambda_1^{(\rm core)}$ is particularly small, for example $1-\lambda_1^{(\rm core)} \sim 10^{-17}$, such that in standard double precision arithmetic the Arnoldi method, applied on the matrix S_{cc}, does not allow to determine this small gap. For these particular cases it is possible to determine rather accurately the core space gap and the corresponding eigenvector by another numerical approach called "projected power method" (Frahm *et al.*, 2011). These eigenvectors, shown in Fig. 19 (b), are strongly localized on a modest number of nodes $\sim 10^2$ and with very small but non-vanishing values on the other nodes. Technically these vectors extend to the whole core space but practically they define small quasi-subspaces (in the core space domain) where the escape probability is extremely small (Frahm *et al.*, 2011) and in the range $1-\alpha \sim 10^{-8}$ they still contribute to the PageRank according to Eq. (9).

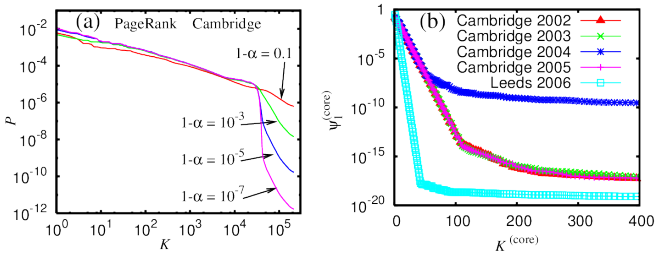

FIG. 19 (Color online) (a) PageRank $P(K)$ of Cambridge 2006 for $1-\alpha = 0.1$, 10^{-3}, 10^{-5}, 10^{-7}. (b) First core space eigenvector $\psi_1^{(\rm core)}$ versus its rank index $K^{(\rm core)}$ for the UK university networks with a small core space gap $1-\lambda_1^{(\rm core)} < 10^{-8}$. After (Frahm *et al.*, 2011).

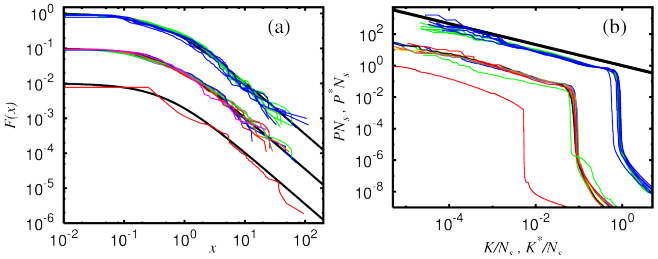

FIG. 20 (Color online) (a) Fraction of invariant subspaces F with dimensions larger than d as a function of the rescaled variable $x = d/\langle d\rangle$. Upper curves correspond to Cambridge (green/gray) and Oxford (blue/black) for years 2002 to 2006 and middle curves (shifted down by a factor of 10) correspond to the university networks of Glasgow, Cambridge, Oxford, Edinburgh, UCL, Manchester, Leeds, Bristol and Birkbeck for year 2006 with $\langle d\rangle$ between 14 and 31. Lower curve (shifted down by a factor of 100) corresponds to the matrix S^* of Wikipedia with $\langle d\rangle = 4$. The thick black line is $F(x) = (1+2x)^{-1.5}$. (b) Rescaled PageRank $P\,N_s$ versus rescaled rank index K/N_s for $1-\alpha = 10^{-8}$ and $3974 \leq N_s \leq 48239$ for the same university networks as in (a) (upper and middle curves, the latter shifted down and left by a factor of 10). The lower curve (shifted down and left by a factor of 100) shows the rescaled CheiRank of Wikipedia $P^*\,N_s$ versus K^*/N_s with $N_s = 21198$. The thick black line corresponds to a power law with exponent $-2/3$. After (Frahm *et al.*, 2011).

In Fig. 20(b) we show that for several of the university networks the PageRank at $1-\alpha = 10^{-8}$ has actually a universal form when using the rescaled variables $P\,N_s$ versus K/N_s with a power law behavior close to $P \propto K^{-2/3}$ for $K/N_s < 1$. The rescaled data of Fig. 20 (a) show that the fraction of subspaces with dimensions larger than d is well described by the power law $F(x) \approx (1+2x)^{-1.5}$ with the dimensionless variable $x = d/\langle d\rangle$ where $\langle d\rangle$ is an average subspace dimension computed for WWW of a given university. The tables of all considered UK universities with the parameters of their WWW are given in (Frahm *et al.*, 2011). We note that the CheiRank of S^* of Wikipedia 2009 also approximately follows the above universal distributions. How-

ever, for S matrix of Wikipedia the number of subspaces is small and statistical analysis cannot be performed for this case.

The origin of the universal distribution $F(x)$ still remains a puzzle. Possible links with a percolation on directed networks (see e.g. (Dorogovtsev *et al.*, 2008)) are still to be elucidated. It also remains unclear how stable this distribution really is. It works well for UK university networks 2002-2006. However, for the Twitter network (Frahm and Shepelyansky , 2012b) such a distribution becomes rather approximate. Also for the network of Cambridge in 2011, analyzed in (Ermann *et al.*, 2012a, 2013b) with $N \approx 8.9 \times 10^5$, $N_\ell \approx 1.5 \times 10^7$, the number of subspaces is significantly reduced and a statistical analysis of their size distribution becomes not relevant. It is possible that an increase of number of links per node N_ℓ/N from a typical value of 10 for UK universities to 35 for Twitter affects this distribution. For Cambridge 2011 the network entered in a regime when many links are generated by robots that apparently leads to a change of its statistical properties.

C. Two-dimensional ranking for University networks

Two-dimensional ranking of network nodes provides a new characterization of directed networks. Here we consider a density distribution of nodes (see Sec. IV.C) in the PageRank-CheiRank plane for examples of two WWW networks of Cambridge 2006 and ENS Paris 2011 shown in Fig. 21 from (Ermann *et al.*, 2012a).

The density distribution for Cambridge 2006 clearly shows that nodes with high PageRank have low CheiRank that corresponds to zero density at low K, K^* values. At large K, K^* values there is a maximum line of density which is located not very far from the diagonal $K \approx K^*$. The presence of correlations between $P(K_i)$ and $P^*(K_i{}^*)$ leads to a probability distribution with one main maximum along a diagonal at $\ln K + \ln K^* = const$. This is similar to the properties of the density distribution for the Wikipedia network shown in Fig. 7(a).

The 2DRanking might give new possibilities for information retrieval from large databases which are growing rapidly with time. Indeed, for example the size of the Cambridge network increased by a factor 4 from 2006 to 2011. At present, web robots start automatically to generate new web pages. These features can be responsible for the appearance of gaps in the density distribution in (K, K^*) plane at large $K, K^* \sim N$ values visible for large scale university networks such as ENS Paris in 2011 (see Fig. 21). Such an automatic generation of links can change the scale-free properties of networks. Indeed, for ENS Paris a large step in the PageRank distribution appears (Ermann *et al.*, 2012a) possibly indicating a delocalization transition tendency of the PageRank that can destroy the efficiency of information retrieval from the WWW.

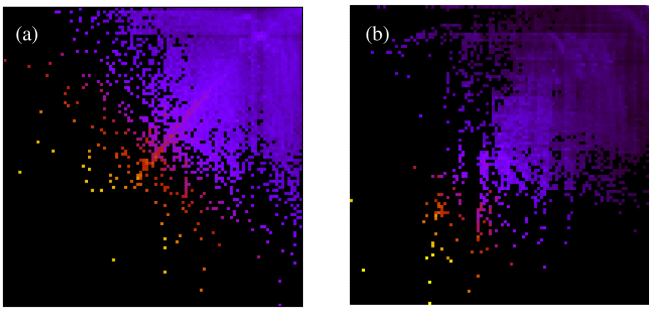

FIG. 21 (Color online) Density distribution $W(K, K^*) = dN_i/dKdK^*$ for networks of Universities in the plane of PageRank K and CheiRank K^* indexes in log-scale $(\log_N K, \log_N K^*)$. The density is shown for 100×100 equidistant grid in $\log_N K, \log_N K^* \in [0, 1]$, the density is averaged over all nodes inside each cell of the grid, the normalization condition is $\sum_{K,K^*} W(K, K^*) = 1$. Color varies from black for zero to yellow/gray for maximum density value W_M with a saturation value of $W_s^{1/4} = 0.5 W_M^{1/4}$ so that the same color is fixed for $0.5 W_M^{1/4} \leq W^{1/4} \leq W_M^{1/4}$ to show in a better way low densities. The panels show networks of University of Cambridge 2006 with $N = 212710$ (a) and ENS Paris 2011 for crawling level 7 with $N = 1820015$ (b). After (Ermann *et al.*, 2012a).

IX. WIKIPEDIA NETWORKS

The free online encyclopedia Wikipedia is a huge repository of human knowledge. Its size is growing permanently accumulating enormous amount of information and becoming a modern version of *Library of Babel*, described by Jorge Luis Borges (Borges, 1962). The hyperlink citations between Wikipedia articles provides an important example of directed networks evolving in time for many different languages. In particular, the English edition of August 2009 has been studied in detail (Ermann *et al.*, 2012a, 2013b; Zhirov *et al.*, 2010). The effects of time evolution (Eom *et al.* , 2013b) and entanglement of cultures in multilingual Wikipedia editions have been investigated in (Aragón *et al.*, 2012; Eom and Shepelyansky , 2013a; Eom *et al.* , 2014).

A. Two-dimensional ranking of Wikipedia articles

The statistical distribution of links in Wikipedia networks has been found to follow a power law with the exponents $\mu_{\rm in}, \mu_{\rm out}$ (see e.g. (Capocci *et al.*, 2006; Muchnik *et al.*, 2007; Zhirov *et al.*, 2010; Zlatic *et al.*, 2006)). The probabilities of PageRank and CheiRank are shown in Fig. 5. They are satisfactory described by a power law decay with exponents $\beta_{PR,CR} = 1/(\mu_{\rm in,out} - 1)$ (Zhirov *et al.*, 2010).

The density distribution of articles over PageRank-CheiRank plane $(\log_N K, \log_N K^*)$ is shown in Fig. 7(a) for English Wikipedia Aug 2009. We stress that the den-

sity is very different from those generated by the product of independent probabilities of P and P^* given in Fig. 5. In the latter case we obtain a density homogeneous along lines $\ln K^* = -\ln K + const$ being rather similar to the distribution for Linux network also shown in Fig. 7. This result is in good agreement with a fact that the correlator κ between PageRank and CheiRank vectors is rather large for Wikipedia $\kappa = 4.08$ while it is close to zero for Linux network $\kappa \approx -0.05$.

The difference between PageRank and CheiRank is clearly seen from the names of articles with highest ranks (ranks of all articles are given in (Zhirov *et al.*, 2010)). At the top of PageRank we have 1. *United States*, 2. *United Kingdom*, 3. *France* while for CheiRank we find 1. *Portal:Contents/Outline of knowledge/Geography and places*, 2. *List of state leaders by year*, 3. *Portal:Contents/Index/Geography and places*. Clearly PageRank selects first articles on a broadly known subject with a large number of ingoing links while CheiRank selects first highly communicative articles with many outgoing links. The 2DRank combines these two characteristics of information flow on directed network. At the top of 2DRank K_2 we find 1. *India*, 2. *Singapore*, 3. *Pakistan*. Thus, these articles are most known/popular and most communicative at the same time.

The top 100 articles in K, K_2, K^* are determined for several categories including countries, universities, people, physicists. It is shown in (Zhirov *et al.*, 2010) that PageRank recovers about 80% of top 100 countries from SJR data base (SJR , 2007), about 75% of top 100 universities of Shanghai university ranking (Shanghai ranking , 2010), and, among physicists, about 50% of top 100 Nobel winners in physics. This overlap is lower for 2DRank and even lower for CheiRank. However, as we will see below in more detail, 2DRank and CheiRank highlight other properties being complementary to PageRank.

Let us give an example of top three physicists among those of 754 registered in Wikipedia in 2010: 1. *Aristotle*, 2. *Albert Einstein*, 3. *Isaac Newton* from PageRank; 1. *Albert Einstein*, 2. *Nikola Tesla*, 3. *Benjamin Franklin* from 2DRank; 1. *Hubert Reeves*, 2. *Shen Kuo*, 3. *Stephen Hawking* from CheiRank. It is clear that PageRank gives most known, 2DRank gives most known and active in other areas, CheiRank gives those who are known and contribute to popularization of science. Indeed, e.g. *Hubert Reeves* and *Stephen Hawking* are very well known for their popularization of physics that increases their communicative power and place them at the top of CheiRank. *Shen Kuo* obtained recognized results in an enormous variety of fields of science that leads to the second top position in CheiRank even if his activity was about thousand years ago.

According to Wikipedia ranking the top universities are 1. *Harvard University*, 2. *University of Oxford*, 3. *University of Cambridge* in PageRank; 1. *Columbia University*, 2. *University of Florida*, 3. *Florida State University* in 2DRank and CheiRank. CheiRank and 2DRank highlight connectivity degree of universities that leads to appearance of significant number of arts, religious and military specialized colleges (12% and 13% respectively for CheiRank and 2DRank) while PageRank has only 1% of them. CheiRank and 2DRank introduce also a larger number of relatively small universities who are keeping links to their alumni in a significantly better way that gives an increase of their ranks. It is established (Eom *et al.* , 2013b) that top 10 PageRank universities from English Wikipedia in years $2003, 2005, 2007, 2009, 2011$ recover correspondingly $9, 9, 8, 7, 7$ from top 10 of (Shanghai ranking , 2010).

The time evolution of probability distributions of PageRank, CheiRank and two-dimensional ranking is analyzed in (Eom *et al.* , 2013b) showing that they become stabilized for the period 2007-2011.

On the basis of these results we can conclude that the above algorithms provide correct and important ranking of huge information and knowledge accumulated at Wikipedia. It is interesting that even Dow-Jones companies are ranked via Wikipedia networks in a good manner (Zhirov *et al.*, 2010). We discuss ranking of top people of Wikipedia a bit later.

B. Spectral properties of Wikipedia network

The complex spectrum of eigenvalues of G for English Wikipedia network of Aug 2009 is shown in Fig. 22. As for university networks, the spectrum also has some invariant subspaces resulting in degeneracies of the leading eigenvalue $\lambda = 1$ of S (or S^*). However, due to the stronger connectivity of the Wikipedia network these subspaces are significantly smaller compared to university networks (Eom *et al.* , 2013b; Ermann *et al.*, 2013b). For example of Aug 2009 edition in Fig. 22 there are 255 invariant subspaces (of the matrix S) covering 515 nodes with 255 unit eigenvalues $\lambda_j = 1$ and 381 eigenvalues on the complex unit circle with $|\lambda_j| = 1$. For the matrix S^* of Wikipedia there are 5355 invariant subspaces with 21198 nodes, 5365 unit eigenvalues and 8968 eigenvalues on the unit circle (Ermann *et al.*, 2013b). The complex spectra of all subspace eigenvalues and the first $n_A = 6000$ core space eigenvalues of S and S^* are shown in Fig. 22. As in the university cases, in the spectrum we can identify cross and triple-star structures similar to those of orthostochastic matrices shown in Fig. 18. However, for Wikipedia (especially for S) the largest complex eigenvalues outside the real axis are more far away from the unit circle. For S of Wikipedia the two largest core space eigenvalues are $\lambda_1^{(\rm core)} = 0.999987$ and $\lambda_2^{(\rm core)} = 0.977237$ indicating that the core space gap $|1 - \lambda_1^{(\rm core)}| \sim 10^{-5}$ is much smaller than the secondary gap $|\lambda_1^{(\rm core)} - \lambda_2^{(\rm core)}| \sim 10^{-2}$. As a consequence the PageRank of Wikipedia (at $\alpha = 0.85$) is strongly influenced by the leading core space eigenvector and actually both vectors select the same 5 top nodes.

The time evolution of spectra of G and G^* for English Wikipedia is studied in (Eom *et al.* , 2013b). It is shown that the spectral structure remains stable for years 2007 - 2011.

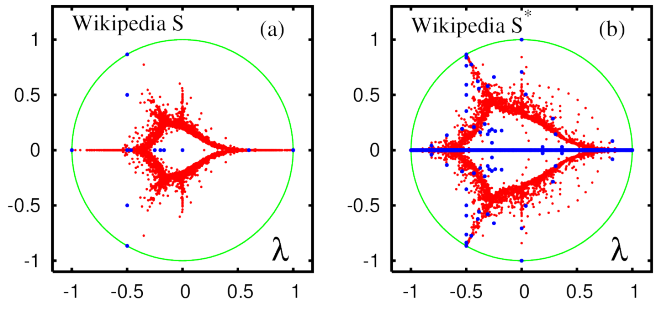

FIG. 22 (Color online) Complex eigenvalue spectra λ of S (a) and S^* (b) for English Wikipedia of Aug 2009 with $N = 3282257$ articles and $N_\ell = 71012307$ links. Red/gray dots are core space eigenvalues, blue/black dots are subspace eigenvalues and the full green/gray curve shows the unit circle. The core space eigenvalues are computed by the projected Arnoldi method with Arnoldi dimension $n_A = 6000$. After (Eom *et al.* , 2013b).

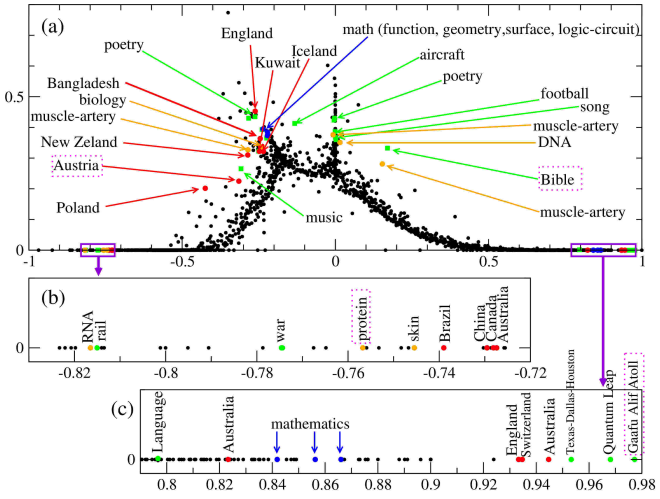

FIG. 23 (Color online) Complex eigenvalue spectrum of the matrices S for English Wikipedia Aug 2009. Highlighted eigenvalues represent different communities of Wikipedia and are labeled by the most repeated and important words following word counting of first 1000 nodes. Panel (a) shows complex plane for positive imaginary part of eigenvalues, while panels (b) and (c) zoom in the negative and positive real parts. After (Ermann *et al.*, 2013b).

C. Communities and eigenstates of Google matrix

The properties of eigenstates of Gogle matrix of Wikipedia Aug 2009 are analyzed in (Ermann *et al.*, 2013b). The global idea is that the eigenstates with large values of $|\lambda|$ select certain specific communities. If $|\lambda|$ is close to unity then a relaxation of probability from such nodes is rather slow and we can expect that such eigenstates highlight some new interesting information even if these nodes are located on a tail of PageRank. The important advantage of the Wikipedia network is that its nodes are Wikipedia articles with a relatively clear meaning allowing to understand the origins of appearance of certain nodes in one community.

The localization properties of eigenvectors ψ_i of the Google matrix can be analyzed with the help of IPR ξ (see Sec. III.E). Another possibility is to fit a decay of an eigenstate amplitude by a power law $|\psi_i(K_i)| \sim K_i^b$ where K_i is the index ordering $|\psi_i(j)|$ by monotonically decreasing amplitude (similar to $P(K)$ for PageRank). The exponents b on the tails of $|\psi_i(j)|$ are found to be typically in the range $-2 < b < -1$ (Ermann *et al.*, 2013b). At the same time the eigenvectors with large complex eigenvalues or real eigenvalues close to ± 1 are quite well localized on $\xi_i \approx 10^2 - 10^3$ nodes that is much smaller than the whole network size $N \approx 3 \times 10^6$.

To understand the meaning of other eigenstates in the core space we order selected eigenstates by their decreasing value $|\psi_i(j)|$ and apply word frequency analysis for the first 1000 articles with $K_i \leq 1000$. The mostly frequent word of a given eigenvector is used to label the eigenvector name. These labels with corresponding eigenvalues are shown in Fig. 23. There are four main categories for the selected eigenvectors belonging to countries (red/gray), biology and medicine (orange/very light gray), mathematics (blue/black) and others (green/light gray). The category of others contains rather diverse articles about poetry, Bible, football, music, American TV series (e.g. Quantum Leap), small geographical places (e.g. Gaafru Alif Atoll). Clearly these eigenstates select certain specific communities which are relatively weakly coupled with the main bulk part of Wikipedia that generates relatively large modulus of $|\lambda_i|$.

For example, for the article *Gaafu Alif Atoll* the eigenvector is mainly localized on names of small atolls forming *Gaafu Alif Atoll*. Clearly this case represents well localized community of articles mainly linked between themselves that gives slow relaxation rate of this eigenmode with $\lambda = 0.9772$ being rather close to unity. Another eigenvector has a complex eigenvalue with $|\lambda| = 0.3733$ and the top article *Portal:Bible*. Another two articles are *Portal:Bible/Featured chapter/archives*, *Portal:Bible/Featured article*. These top 3 articles have very close values of $|\psi_i(j)|$ that seems to be the reason why we have $\varphi = \arg(\lambda_i) = 0.3496\pi$ being very close to $\pi/3$. Examples of other eigenvectors are discussed in (Ermann *et al.*, 2013b) in detail.

The analysis performed in (Ermann *et al.*, 2013b) for Wikipedia Aug 2009 shows that the eigenvectors of the Google matrix of Wikipedia clearly identify certain communities which are relatively weakly connected with the Wikipedia core when the modulus of corresponding eigenvalue is close to unity. For moderate values of $|\lambda|$

we still have well defined communities which are however have stronger links with some popular articles (e.g. countries) that leads to a more rapid decay of such eigenmodes. Thus the eigenvectors highlight interesting features of communities and network structure. However, a priori, it is not evident what is a correspondence between the numerically obtained eigenvectors and the specific community features in which someone has a specific interest. In fact, practically each eigenvector with a moderate value $|\lambda| \sim 0.5$ selects a certain community and there are many of them. So it remains difficult to target and select from eigenvalues λ a specific community one is interested.

The spectra and eigenstates of other networks like WWW of Cambridge 2011, Le Monde, BBC and PCN of Python are discussed in (Ermann *et al.*, 2013b). It is found that IPR values of eigenstates with large $|\lambda|$ are well localized with $\xi \ll N$. The spectra of each network have significant differences from one another.

D. Top people of Wikipedia

There is always a significant public interest to know who are most significant historical figures, or persons, of humanity. The Hart list of the top 100 people who, according to him, most influenced human history, is available at (Hart, 1992). Hart "ranked these 100 persons in order of importance: that is, according to the total amount of influence that each of them had on human history and on the everyday lives of other human beings" (Hart, 1992). Of course, a human ranking can be always objected arguing that an investigator has its own preferences. Also investigators from different cultures can have different view points on a same historical figure. Thus it is important to perform ranking of historical figures on purely mathematical and statistical grounds which exclude any cultural and personal preferences of investigators.

A detailed two-dimensional ranking of persons of English Wikipedia Aug 2009 has been done in (Zhirov *et al.*, 2010). Earlier studies had been done in a non-systematic way without any comparison with established top 100 lists (see these Refs. in (Wikipedia Top 100, 2014; Zhirov *et al.*, 2010)). Also at those times Wikipedia did not yet entered in its stabilized phase of development.

The top people of Wikipedia Aug 2009 are found to be 1. *Napoleon I of France*, 2. *George W. Bush*, 3. *Elizabeth II of the United Kingdom* for PageRank; 1.*Michael Jackson*, 2. *Frank Lloyd Wright*, 3. *David Bowie* for 2DRank; 1. *Kasey S. Pipes*, 2. *Roger Calmel*, 3. *Yury G. Chernavsky* for CheiRank (Zhirov *et al.*, 2010). For the PageRank list of 100 the overlap with the Hart list is at 35% (PageRank), 10% (2DRank) and almost zero for CheiRank. This is attributed to a very broad distribution of historical figures on 2D plane, as shown in Fig. 7, and a large variety of human activities. These activities are classified by 5 main categories: politics, religion, arts, science, sport. For the top 100 PageRank persons we have the following distribution over these categories: 58, 10, 17, 15, 0 respectively. Clearly PageRank overestimates the significance of politicians which list is dominated by USA presidents not always much known to a broad public. For 2DRank we find respectively 24, 5, 62, 7, 2. Thus this rank highlights artistic sides of human activity. For CheiRank we have 15, 1, 52, 16, 16 so that the dominant contribution comes from arts, science and sport. The interesting property of this rank is that it selects many composers, singers, writers, actors. As an interesting feature of CheiRank we note that among scientists it selects those who are not so much known to a broad public but who discovered new objects, e.g. George Lyell who discovered many Australian butterflies or Nikolai Chernykh who discovered many asteroids. CheiRank also selects persons active in several categories of human activity.

For English Wikipedia Aug 2009 the distribution of top 100 PageRank, CheiRank and Hart's persons on PageRank-CheiRank plane is shown in Fig. 7 (a).

The distribution of Hart's top 100 persons on (K, K^*) plane for English Wikipedia in years 2003, 2005, 2007, Aug 2009, Dec 2009, 2011 is found to be stable for the period 2007-2011 even if certain persons change their ranks (Eom *et al.* , 2013b). The distribution of top 100 persons of Wikipedia Aug 2009 remains stable and compact for PageRank and 2DRank for the period 2007-2011 while for CheiRank the fluctuations of positions are large. This is due to the fact that outgoing links are easily modified and fluctuating.

The time evolution of distribution of top persons over fields of human activity is established in (Eom *et al.* , 2013b). PageRank persons are dominated by politicians whose percentage increases with time, while the percent of arts decreases. For 2DRank the arts are dominant but their percentage decreases with time. We also see the appearance of sport which is absent in PageRank. The mechanism of the qualitative ranking differences between two ranks is related to the fact that 2DRank takes into account via CheiRank a contribution of outgoing links. Due to that singers, actors, sportsmen improve their CheiRank and 2DRrank positions since articles about them contain various music albums, movies and sport competitions with many outgoing links. Due to that the component of arts gets higher positions in 2DRank in contrast to dominance of politics in PageRank.

The interest to ranking of people via Wikipedia network is growing, as shows the recent study of English edition (Skiena and Ward, 2014).

E. Multilingual Wikipedia editions

The English edition allows to obtain ranking of historical people but as we saw the PageRank list is dominated by USA presidents that probably does not correspond to the global world view point. Hence, it is important to study multilingual Wikipedia editions which have now

287 languages and represent broader cultural views of the world.

One of the first cross-cultural study was done for 15 largest language editions constructing a network of links between set of articles of people biographies for each edition. However, the number of nodes and links in such a biographical network is significantly smaller compared to the whole network of Wikipedia articles and thus the fluctuations become rather large. For example, from the biographical network of the Russian edition one finds as the top person *Napoleon III* (and even not *Napoleon I*) (Aragón *et al.*, 2012), who has a rather low importance for Russia.

Another approach was used in (Eom and Shepelyansky , 2013a) ranking top 30 persons by PageRank, 2DRank and CheiRank algorithms for all articles of each of 9 editions and attributing each person to her/his native language. The selected editions are English (EN), French (FR), German (DE), Italian (IT), Spanish (ES), Dutch (NL), Russian (RU), Hungarian (HU) and Korean (KO). The aim here is to understand how different cultures evaluate a person? Is an important person in one culture is also important in the other culture? It is found that local heroes are dominant but also global heroes exist and create an effective network representing entanglement of cultures.

The top article of PageRank is usually *USA* or the name of country of a given language (FR, RU, KO). For NL we have at the top *beetle, species, France*. The top articles of CheiRank are various listings.

The distributions of articles density and top 30 persons for each rank algorithm are shown in Fig. 24 for four editions EN, FR, DE, RU. We see that in global the distributions have a similar shape that can be attributed to a fact that all editions describe the same world. However, local features of distributions are different corresponding to different cultural views on the same world (other 5 editions are shown in Fig.2 in (Eom and Shepelyansky , 2013a)). The top 30 persons for each edition are selected manually that represents a weak point of this study.

From the lists of top persons, the "fields" of activity are identified for each top 30 rank persons in which he/she is active on. The six activity fields are: politics, art, science, religion, sport and etc (here "etc" includes all other activities). As shown in Fig. 25, for PageRank, politics is dominant and science is secondarily dominant. The only exception is Dutch where science is the almost dominant activity field (politics has the same number of points). In case of 2DRank in Fig. 25, art becomes dominant and politics is secondarily dominant. In case of CheiRank, art and sport are dominant fields (see Fig.3 in (Eom and Shepelyansky , 2013a)). Thus for example, in CheiRank top 30 list we find astronomers who discovered a lot of asteroids, e.g. Karl Wilhelm Reinmuth (4th position in RU and 7th in DE), who was a prolific discoverer of about 400 of them. As a result, his article contains a long listing of asteroids discovered by him and giving him a high CheiRank. The distributions of persons over activity fields are shown in Fig. 25 for 9 languages editions (marked by standard two letters used by Wikipedia).

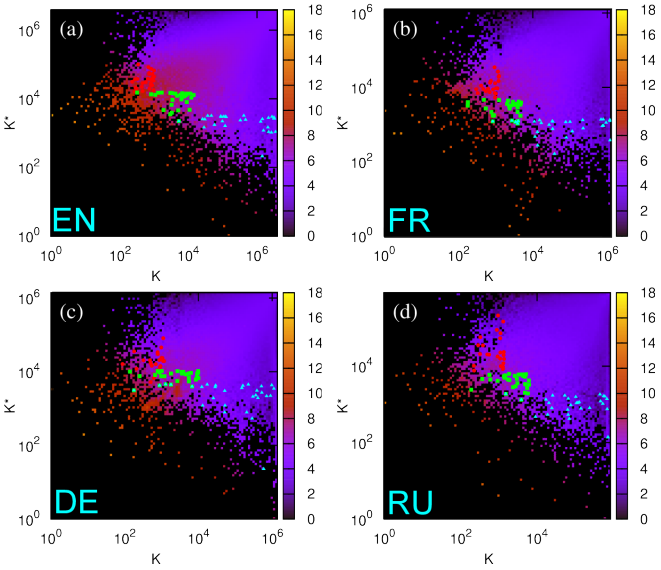

FIG. 24 (Color online) Density of Wikipedia articles in the PageRank-CheiRank plane (K, K^*) for four different language Wikipedia editions. The red (gray) points are top PageRank articles of persons, the green (light gray) squares are top 2DRank articles of persons and the cyan (dark gray) triangles are top CheiRank articles of persons. Wikipedia language editions are English EN (a), French FR (b), German DE (c), and Russian RU (d). Color bars show natural logarithm of density, changing from minimal nonzero density (dark) to maximal one (white), zero density is shown by black. After (Eom and Shepelyansky , 2013a).

The change of activity priority for different ranks is due to the different balance between incoming and outgoing links there. Usually the politicians are well known for a broad public, hence, the articles about politicians are pointed by many articles. However, the articles about politicians are not very communicative since they rarely point to other articles. In contrast, articles about persons in other fields like science, art and sport are more communicative because of listings of insects, planets, asteroids they discovered, or listings of song albums or sport competitions they gain.

On the basis of this approach one obtains local ranks of each of 30 persons $1 \leq K_{P,E,A} \leq 30$ for each edition E and algorithm A. Then an average ranking score of a person P is determined as $\Theta_{P,A} = \sum_E (31 - K_{P,E,A})$ for each algorithm. This method determines the global historical figures. The top global persons are 1.*Napoleon*, 2.*Jesus*, 3.*Carl Linnaeus* for PageRank; 1.*Micheal Jackson* , 2.*Adolf Hitler*, 3.*Julius Caesar* for 2DRank. For CheiRank the lists of different editions have rather low overlap and such an averaging is not efficient. The first positions reproduce top persons from English edition discussed in Sec. IX.D, however, the next ones are different.

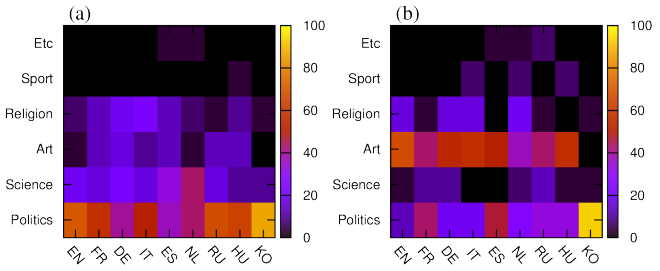

FIG. 25 (Color online) Distribution of top 30 persons over activity fields for PageRank (a) and 2DRank (b) for each of 9 Wikipedia editions. The color bar shows the values in percent. After (Eom and Shepelyansky , 2013a).

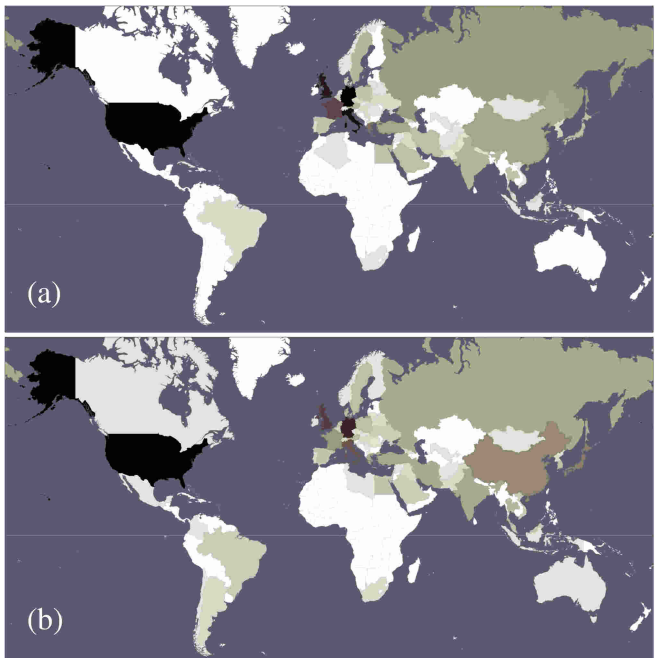

FIG. 26 Number of appearances of historical figures of a given country, obtained from 24 lists of top 100 persons of PageRank (a) and 2DRank (b), shown on the world map. Color changes from zero (white) to maximum (black), it corresponds to average number of person appearances per country. After (Eom *et al.* , 2014).

Since each person is attributed to her/his native language it is also possible for each edition to obtain top local heroes who have native language of the edition. For example, we find for PageRank for EN *George W. Bush*, *Barack Obama*, *Elizabeth II*; for FR *Napoleon*, *Louis XIV of France*, *Charles de Gaulle*; for DE *Adolf Hitler*, *Martin Luther*, *Immanuel Kant*; for RU *Peter the Great*, *Joseph Stalin*, *Alexander Pushkin*. For 2DRank we have for EN *Frank Sinatra*, *Paul McCartney*, *Michael Jackson*; for FR *Francois Mitterrand*, *Jacques Chirac*, *Honore de Balzac*; for DE *Adolf Hitler*, *Otto von Bismarck*, *Ludwig van Beethoven*; for RU *Dmitri Mendeleev*, *Peter the Great*, *Yaroslav the Wise*. These ranking results are rather reasonable for each language. Results for other editions and CheiRank are given in (Eom and Shepelyansky , 2013a).

A weak point of above study is a manual selection of persons and a not very large number of editions. A significant improvement has been reached in a recent study (Eom *et al.* , 2014) where 24 editions have been analyzed. These 24 languages cover 59 percent of world population, and these 24 editions covers 68 percent of the total number of Wikipedia articles in all 287 available languages. Also the selection of people from the rank list of each edition is now done in an automatic computerized way. For that a list of about 1.1 million biographical articles about people with their English names is generated. From this list of persons, with their biographical article title in the English Wikipedia, the corresponding titles in other language editions are determined using the inter-language links provided by Wikipedia.

Using the corresponding articles, identified by the inter-languages links in different language editions, the top 100 persons are obtained from the rankings of all Wikipedia articles of each edition. A birth place, birth date, and gender of each top 100 ranked person are identified, based on DBpedia or a manual inspection of the corresponding Wikipedia biographical article, when for the considered person no DBpedia data were available. In this way 24 lists of top 100 persons for each edition are obtained in PageRank with 1045 unique names and in 2DRank with 1616 unique names. Each of the 100 historical figures is attributed to a birth place at the country level, to a birth date in year, to a gender, and to a cultural language group. The birth place is assigned according to the current country borders. The cultural group of historical figures is assigned by the most spoken language of their birth place at the current country level. The considered editions are: English EN, Dutch NL, German DE, French FR, Spanish, ES, Italian IT, Potuguese PT, Greek, EL, Danish DA, Swedish SV, Polish PL, Hungarian HU, Russian RU, Hebrew HE, Turkish TR, Arabic AR, Persian FA, Hindi HI, Malaysian MS, Thai TH, Vietnamese VI, Chinese ZH, Korean KO, Japanese JA (dated by February 2013). The size of network changes from maximal value $N = 4212493$ for EN to minimal one $N = 78953$ for TH.

All persons are ranked by their average rank score $\Theta_{P,A} = \sum_E (101 - K_{P,E,A})$ with $1 \leq K_{P,E,A} \leq 100$ similar to the study of 9 editions described above. For PageRank the top global historical figures are *Carl Linnaeus*, *Jesus*, *Aristotle* and for 2DRank we obtain *Adolf Hitler*, *Michael Jackson*, *Madonna (entertainer)*. Thus the averaging over 24 editions modifies the top ranking. The list of top 100 PageRank global persons has overlap of 43 persons with the Hart list (Hart, 1992). Thus the averaging over 24 editions gives a significant improvement compared to 35 persons overlap for the case of English edition only (Zhirov *et al.*, 2010). For comparison we note that the top 100 list of historical figures has been also determined recently by (Pantheon MIT project, 2014) having

overlap of 42 persons with the Hart list. This Pantheon MIT list is established on the basis of number of editions and number of clicks on an article of a given person without using rank algorithms discussed here. The overlap between top 100 PageRank list and top 100 Pantheon list is 44 percent. More data are available in (Eom *et al.* , 2014).

The fact that *Carl Linnaeus* is the top historical figure of Wikipedia PageRank list came out as a surprise for media and broad public (see (Wikipedia Top 100, 2014)). This ranking is due to the fact that *Carl Linnaeus* created a classification of world species including, animals, insects, herbs, trees etc. Thus all articles of these species point to the article *Carl Linnaeus* in various languages. As a result *Carl Linnaeus* appears on almost top positions in all 24 languages. Hence, even if a politician, like *Barak Obama*, takes the second position in his country language EN (*Napoleon* is at the first position in EN) he is usually placed at low ranking in other language editions. As a result *Carl Linnaeus* takes the first global PageRank position.

The number of appearances of historical persons in 24 lists of top 100 for each edition can be distributed over present world countries according to the birth place of each person. This geographical distribution is shown in Fig. 26 for PageRank and 2DRank. In PageRank the top countries are *DE, USA, IT* and in 2DRank *US, DE, UK*. The appearance of many UK and US singers improves the positions of English speaking countries in 2DRank.

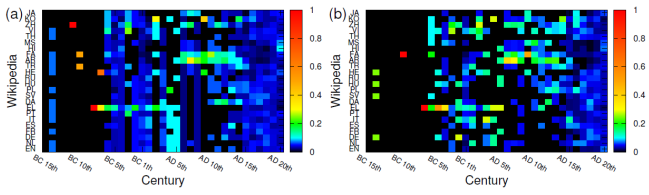

FIG. 27 (Color online) Birth date distributions over 35 centuries of top historical figures from each Wikipedia edition marked by two letters standard notation of Wikipedia. Panels: (a) column normalized birth date distributions of PageRank historical figures; (b) same as (a) for 2DRank historical figures. After (Eom *et al.* , 2014).

The distributions of the top PageRank and 2DRank historical figures over 24 Wikipedia editions for each century are shown in Fig. 27. Each person is attributed to a century according to the birth date covering the range of 35 centuries from BC 15th to AD 20th centuries. For each century the number of persons for each century is normalized to unity to see more clearly relative contribution of each language for each century.

The Greek edition has more historical figures in BC 5th century because of Greek philosophers. Also most of western-southern European language editions, including English, Dutch, German, French, Spanish, Italian, Portuguese, and Greek, have more top historical figures because they have Augustine the Hippo and Justinian I in common. The Persian (FA) and the Arabic (AR) Wikipedia have more historical figures comparing to other language editions (in particular European language editions) from the 6th to the 12th century that is due to Islamic leaders and scholars. The data of Fig. 27 clearly show well pronounced patterns, corresponding to strong interactions between cultures: from BC 5th century to AD 15th century for JA, KO, ZH, VI; from AD 6th century to AD 12th century for FA, AR; and a common birth pattern in EN,EL,PT,IT,ES,DE,NL (Western European languages) from BC 5th century to AD 6th century. A detailed analysis shows that even in BC 20th century each edition has a significant fraction of persons of its own language so that even with on going globalization there is a significant dominance of local historical figures for certain cultures. More data on the above points and gender distributions are available in (Eom *et al.* , 2014).

F. Networks and entanglement of cultures

We now know how a person of a given language is ranked by editions of other languages. Therefore, if a top person from a language edition A appears in another edition B, we can consider this as a 'cultural' influence from culture A to B. This generates entanglement in a network of cultures. Here we associate a language edition with its corresponding culture considering that a language is a first element of culture, even if a culture is not reduced only to a language. In (Eom and Shepelyansky , 2013a) a person is attributed to a given language, or culture, according to her/his native language fixed via corresponding Wikipedia article. In (Eom *et al.* , 2014) the attribution to a culture is done via a birth place of a person, each language is considered as a proxy for a cultural group and a person is assigned to one of these cultural groups based on the most spoken language of her/his birth place at the country level. If a person does not belong to any of studied editions then he/she is attributed to an additional cultural group world WR.

After such an attributions of all persons the two networks of cultures are constructed based on the top PageRank historical figures and top 2DRank historical figures respectively. Each culture (i.e. language) is represented as a node of the network, and the weight of a directed link from culture A to culture B is given by the number of historical figures belonging to culture B (e.g. French) appearing in the list of top 100 historical figures for a given culture A (e.g. English).

For example, according to (Eom *et al.* , 2014), there are 5 French historical figures among the top 100 PageRank historical figures of the English Wikipedia, so we can assign weight 5 to the link from English to French. Thus, Fig. 28(a) and Fig. 28(b) represent the constructed networks of cultures defined by appearances of the top PageRank historical figures and top 2DRank historical figures, respectively.

In total we have two networks with 25 nodes which include our 24 editions and an additional node WR for all other world cultures. Persons of a given culture are not taken into account in the rank list of language edition of this culture. Then following the standard rules (1) the Google matrix of network of cultures is constructed by normalization of sum of all elements in each column to unity. The matrix $G_{KK'}$, written in the PageRank indexes K, K' is shown in Fig. 29 for persons from PageRank (a) and 2DRank (b) lists. The matrix G^* is constructed in the same way as G for the network with inverted directions of links.

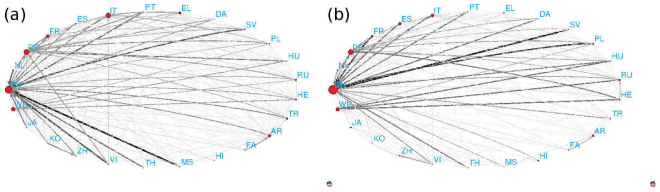

FIG. 28 (Color online) Network of cultures, obtained from 24 Wikipedia languages and the remaining world (WR), considering (a) top 100 PageRank historical figures and (b) top 100 2DRank historical figures. The link width and darkness are proportional to a number of foreign historical figures quoted in top 100 of a given culture, the link direction goes from a given culture to cultures of quoted foreign historical figures, quotations inside cultures are not considered. The size of nodes is proportional to their PageRank. After (Eom *et al.* , 2014).

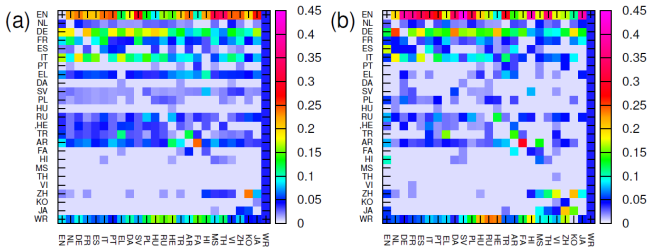

FIG. 29 (Color online) Google matrix of network of cultures shown in Fig 28 (a) and (b) respectively. The matrix elements G_{ij} are shown by color with damping factor $\alpha = 0.85$. After (Eom *et al.* , 2014).

From the obtained matrix G and G^* we determine PageRank and CheiRank vectors and then the PageRank-CheiRank plane (K, K^*), shown in Fig. 30, for networks of cultures from Fig. 28. Here K indicates the ranking of a given culture ordered by how many of its own top historical figures appear in other Wikipedia editions, and K^* indicates the ranking of a given culture according to how many of the top historical figures in the considered culture are from other cultures. It is important to note that for 24 editions the world node WR appears on positions $K = 3$ or $K = 4$, for panels $(a), (b)$ in Fig. 30, signifying that the 24 editions capture the main part of historical figures born in these cultures. We note that for 9 editions in (Eom and Shepelyansky , 2013a) the node WR was at the top position for PageRank so that a significant fraction of historical figures was attributed to other cultures.

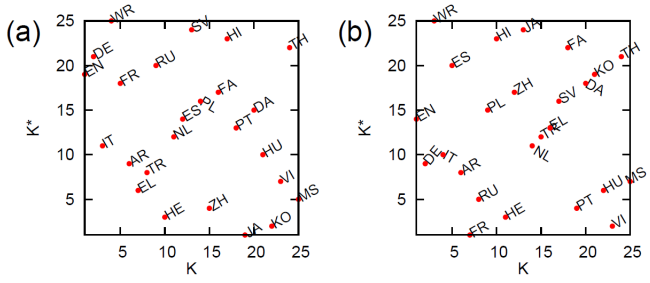

FIG. 30 (Color online) PageRank-CheiRank plane of cultures with corresponding indexes K and K^* obtained from the network of cultures based on (a) top 100 PageRank historical figures, (b) top 100 2DRank historical figures. After (Eom *et al.* , 2014).

From the data of Fig. 30 we obtain at the top positions of K cultures EN, DE, IT showing that other cultures strongly point to them. However, we can argue that for cultures it is also important to have strong communicative property and hence it is important to have 2DRank of cultures at top positions. On the top 2DRank position we have Greek, Turkish and Arabic (for PageRank persons) in Fig. 30(a) and French, Russian and Arabic (for 2DRank persons) in Fig. 30(b). This demonstrates the important historical influence of these cultures both via importance (incoming links) and communicative (outgoing links) properties present in a balanced manner.

Thus the described research across Wikipedia language editions suggests a rigorous mathematical way, based on Markov chains and Google matrix, for recognition of important historical figures and analysis of interactions of cultures at different historical periods and in different world regions. Such an approach recovers 43 percent of persons from the well established Hart historical study (Hart, 1992), that demonstrates the reliability of this method. We think that a further extension of this approach to a larger number of Wikipedia editions will provide a more detailed and balanced analysis of interactions of world cultures.

X. GOOGLE MATRIX OF SOCIAL NETWORKS

Social networks like Facebook, LiveJournal, Twitter, Vkontakte start to play a more and more important role in modern society. The Twitter network is a directed one and here we consider its spectral properties following mainly the analysis reported in (Frahm and Shepelyansky , 2012b).

A. Twitter network

Twitter is a rapidly growing online directed social network. For July 2009 a data set of this entire network is available with $N = 41652230$ nodes and $N_\ell = 1468365182$ links (for data sets see Refs. in (Frahm and Shepelyansky , 2012b)). For this case the spectrum and eigenstate properties of the corresponding Google matrix have been analyzed in detail using the Arnoldi method and standard PageRank and CheiRank computations (Frahm and Shepelyansky , 2012b). For the Twitter network the average number of links per node $\zeta = N_\ell/N \approx 35$ and the general inter-connectivity between top PageRank nodes are considerably larger than for other networks such as Wikipedia (Sec. IX) or UK universities (Sec. VIII) as can be seen in Figs. 31 and 32.

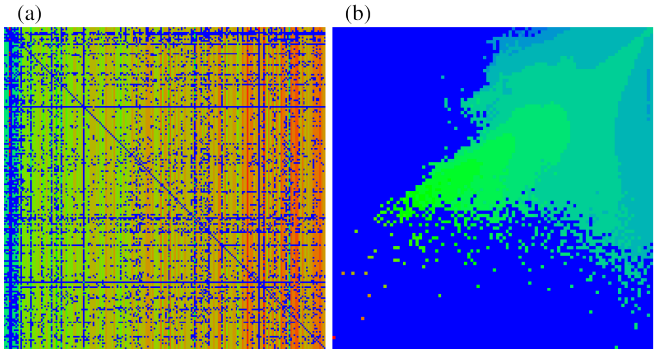

FIG. 31 (Color online) Panel (a): Google matrix of Twitter, matrix elements of G are shown in the basis of PageRank index K of matrix $G_{KK'}$. Here, x (and y) axis show K (and $K^{'}$) with the range $1 \le K, K' \le 200$. Panel (b) shows the density of nodes $W(K, K^*)$ of Twitter on PageRank-CheiRank plane (K, K^*), averaged over 100×100 logarithmically equidistant grids for $0 \le \ln K, \ln K^* \le \ln N$ with the normalization condition $\sum_{K,K^*} W(K, K^*) = 1$. The x-axis corresponds to $\ln K$ and the y-axis to $\ln K^*$. In both panels color varies from blue/black at minimal value to red/gray at maximal value; here $\alpha = 0.85$. After (Frahm and Shepelyansky , 2012b).

The decay of PageRank probability can be approximately described by an algebraic decay with the exponent $\beta \approx 0.54$ while for CheiRank we have a larger value $\beta \approx 0.86$ (Frahm and Shepelyansky , 2012b) that is opposite to the usual situation. The image of top matrix elements of $G_{KK'}$ with $1 \le K, K; \le 200$ is shown in Fig. 31. The density distribution of nodes on (K, K^*) plane is also shown there. It is somewhat similar to those of Wikipedia case in Fig. 24, may be with a larger density concentration along the line $K \approx K^*$.

However, the most striking feature of G matrix elements is a very strong inteconnectivity between top PageRank nodes. Thus for Twitter the top $K \le 1000$ elements fill about 70% of the matrix and about 20% for size $K \le 10^4$. For Wikipedia the filling factor is smaller by a factor $10 - 20$. In particular the number N_G of links between K top PageRank nodes behaves for $K \le 10^3$ as $N_G \sim K^{1.993}$ while for Wikipedia $N_G \sim K^{1.469}$. The exponent for N_G, being close to 2 for Twitter, indicates that for the top PageRank nodes the Google matrix is macroscopically filled with a fraction $0.6 - 0.8$ of non-vanishing matrix elements (see also Figs. 31 and 32) and the very well connected top PageRank nodes can be considered as the Twitter elite (Kandiah and Shepelyansky , 2012). For Wikipedia the interconnectivity among top PageRank nodes has an exponent 1.5 being somewhat reduced but still stronger as compared to certain university networks where typical exponents are close to unity (for the range $10^2 \le K \le 10^4$). The strong interconnectivity of Twitter is also visible in its global logarithmic density distribution of nodes in the PageRank-CheiRank plane (K, K^*) (Fig. 31 (b)) which shows a maximal density along a certain ridge along a line $\ln K^* = \ln K + \text{const}$. with a significant large number of nodes at small values $K, K^* < 1000$.

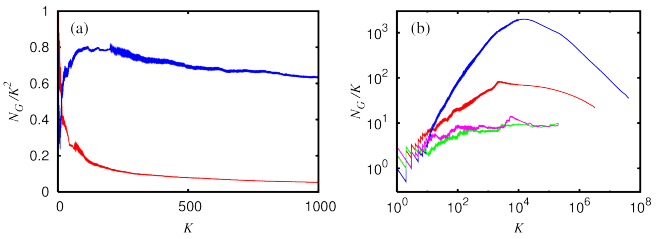

FIG. 32 (Color online) (a) Dependence of the area density $g_K = N_G/K^2$ of nonzero elements of the adjacency matrix among top PageRank nodes on the PageRank index K for Twitter (blue/black curve) and Wikipedia (red/gray curve) networks, data are shown in linear scale. (b) Linear density N_G/K of the same matrix elements shown for the whole range of K in log-log scale for Twitter (blue curve), Wikipedia (red curve), Oxford University 2006 (magenta curve) and Cambridge University 2006 (green curve) (curves from top to bottom at $K = 100$). After (Frahm and Shepelyansky , 2012b).

The decay exponent of the PageRank is for Twitter $\beta = 0.540$ (for $1 \le K \le 10^6$), which indicates a precursor of a delocalization transition as compared to Wikipedia ($\beta = 0.767$) or WWW ($\beta \approx 0.9$), caused by the strong interconnectivity (Frahm and Shepelyansky , 2012b). The Twitter network is also characterized by a large value of PageRank-CheiRank correlator $\kappa = 112.6$ that is by a factor $30 - 60$ larger than this value for Wikipedia and University networks. Such a larger value of κ results from certain individual large values $\kappa_i = NP(K(i))P^*(K^*(i)) \sim 1$. It is argued that this is related to a very strong inter-connectivity between top K PageRank users of the Twitter network (Frahm and Shepelyansky , 2012b).

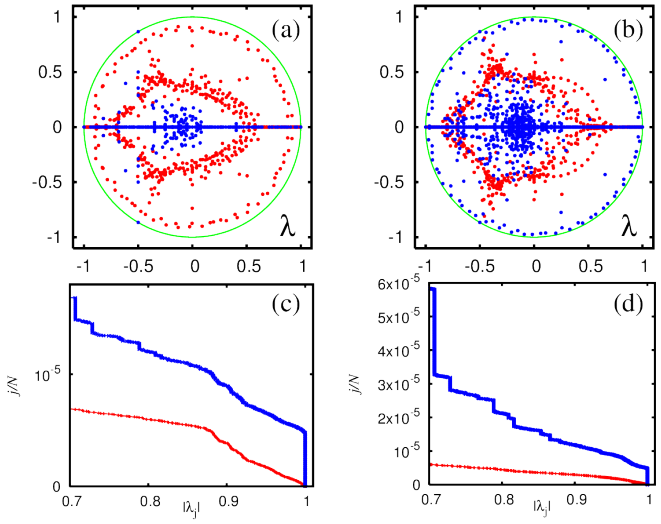

FIG. 33 (Color online) Spectrum of the Twitter matrix S (a) and (c), and S^* (b) and (d). Panels (a) and (b) show subspace eigenvalues (blue/black dots) and core space eigenvalues (red/gray dots) in λ-plane (green/gray curve shows unit circle); there are 17504 (66316) invariant subspaces, with maximal dimension 44 (2959) and the sum of all subspace dimensions is $N_s = 40307$ (180414). The core space eigenvalues are obtained from the Arnoldi method applied to the core space subblock S_{cc} of S with Arnoldi dimension $n_A = 640$. Panels (c) and (d) show the fraction j/N of eigenvalues with $|\lambda| > |\lambda_j|$ for the core space eigenvalues (red/gray bottom curve) and all eigenvalues (blue/black top curve) from raw data ((a) and (b) respectively). The number of eigenvalues with $|\lambda_j| = 1$ is 34135 (129185) of which 17505 (66357) are at $\lambda_j = 1$; this number is (slightly) larger than the number of invariant subspaces which have each at least one unit eigenvalue. Note that in panels (c) and (d) the number of eigenvalues with $|\lambda_j| = 1$ is artificially reduced to 200 in order to have a better scale on the vertical axis. The correct numbers of those eigenvalues correspond to $j/N = 8.195 \times 10^{-4}$ (c) and 3.102×10^{-3} (d) which are strongly outside the vertical panel scale. After (Frahm and Shepelyansky , 2012b).

The spectra of matrices S and S^* are obtained with the help of the Arnoldi method for a relatively modest Arnoldi dimension due to a very large matrix size. The largest n_A modulus eigenvalues $|\lambda|$ are shown in Fig. 33. The invariant subspaces (see Sec. III.C) for the Twitter network cover about $N_s = 4 \times 10^4$ (1.8×10^5) nodes for S (S^*) leading to 1.7×10^4 (6.6×10^4) eigenvalues with $\lambda_j = 1$ or even 3.4×10^4 (1.3×10^5) eigenvalues with $|\lambda_j| = 1$. However, for Twitter the fraction of subspace nodes $g_1 = N_s/N \approx 10^{-3}$ is smaller than the fraction $g_1 \approx 0.2$ for the university networks of Cambridge or Oxford (with $N \approx 2 \times 10^5$) since the size of the whole Twitter network is significantly larger. The complex spectra of S and S^* also show the cross and triple-star structures, as in the cases of Cambridge and Oxford 2006 (see Fig. 17), even though for the Twitter network they are significantly less pronounced.

B. Poisson statistics of PageRank probabilities

From a physical viewpoint one can conjecture that the PageRank probabilities are described by a steady-state quantum Gibbs distribution over certain quantum levels with energies E_i by the identification $P(i) = \exp(-E_i/T)/Z$ with $Z = \sum_i \exp(-E_i/T)$ (Frahm and Shepelyansky , 2014a). In some sense this conjecture assumes that the operator matrix G can be represented as a sum of two operators G_H and G_{NH} where G_H describes a Hermitian system while G_{NH} represents a non-Hermitian operator which creates a system thermalization at a certain effective temperature T with the quantum Gibbs distribution over energy levels E_i of the operator G_H.

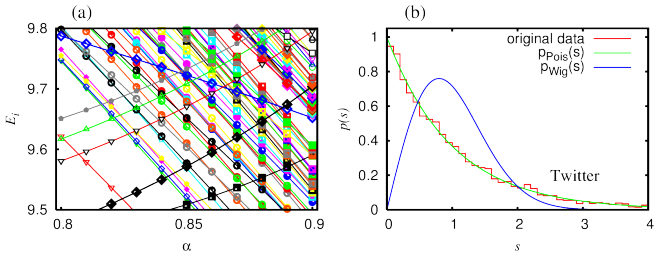

FIG. 34 (Color online) Panel (a) shows the dependence of certain top PageRank levels $E_i = -\ln(P_i)$ on the damping factor α for Twitter network. Data points on curves with one color corresponds to the same node i; about 150 levels are shown close to the minimal energy $E \approx 7.5$. Panel (b) represents the histogram of unfolded level spacing statistics for Twitter at $10 < K \leq 10^4$. The Poisson distribution $p_{\rm Pois}(s) = \exp(-s)$ and the Wigner surmise $p_{\rm Wig}(s) = \frac{\pi}{2}\, s \exp(-\frac{\pi}{4}\, s^2)$ are also shown for comparison. After (Frahm and Shepelyansky , 2014a).

The identification of PageRank with an energy spectrum allows to study the corresponding level statistics which represents a well known concept in the framework of Random Matrix Theory (Guhr *et al.*, 1998; Mehta, 2004). The most direct characteristic is the probability distribution $p(s)$ of unfolded level spacings s. Here $s = (E_{i+1} - E_i)/\Delta E$ is a spacing between nearest levels measured in the units of average local energy spacing ΔE. The unfolding procedure (Guhr *et al.*, 1998; Mehta, 2004) requires the smoothed dependence of E_i on the index K which is obtained from a polynomial fit of $E_i \sim \ln(P_i)$ with $\ln(K)$ as argument (Frahm and Shepelyansky , 2014a).

The statistical properties of fluctuations of levels have been extensively studied in the fields of RMT (Mehta, 2004), quantum chaos (Haake, 2010) and disordered solid state systems (Evers and Mirlin , 2008). It is known that integrable quantum systems have $p(s)$ well described by the Poisson distribution $p_{\rm Pois}(s) = \exp(-s)$. In contrast the quantum systems, which are chaotic in the classical limit (e.g. Sinai billiard), have $p(s)$ given by the RMT being close to the Wigner surmise $p_{\rm Wig}(s) = \frac{\pi}{2}\, s \exp(-\frac{\pi}{4}\, s^2)$ (Bohigas *et al.*, 1984). Also the Ander-

son localized phase is characterized by $p_{\rm Pois}(s)$ while in the delocalized regime one has $p_{\rm Wig}(s)$ (Evers and Mirlin , 2008).

The results for the Twitter PageRank level statistics (Frahm and Shepelyansky , 2014a) are shown in Fig. 34. We find that $p(s)$ is well described by the Poisson distribution. Furthermore, the evolution of energy levels E_i with the variation of the damping factor α shows many level crossings which are typical for Poisson statistics. We may note that here each level has its own index so that it is rather easy to see if there is a real or avoided level crossing.

The validity of the Poisson statistics for PageRank probabilities is confirmed also for the networks of Wikipedia editions in English, French and German from Fig. 24 (Frahm and Shepelyansky , 2014a). We argue that due to absence of level repulsion the PageRank order of nearby nodes can be easily interchanged. The obtained Poisson law implies that the nearby PageRank probabilities fluctuate as random independent variables.

XI. GOOGLE MATRIX ANALYSIS OF WORLD TRADE

During the last decades the trade between countries has been developed in an extraordinary way. Usually countries are ranked in the world trade network (WTN) taking into account their exports and imports measured in *USD* (CIA, 2009). However, the use of these quantities, which are local in the sense that countries know their total imports and exports, could hide the information of the centrality role that a country plays in this complex network. In this section we present the two-dimensional Google matrix analysis of the WTN introduced in (Ermann and Shepelyansky , 2011b). Some previous studies of global network characteristics were considered in (Garlaschelli and Loffredo , 2005; Serrano *et al.*, 2007), degree centrality measures were analyzed in (De Benedictis and Tajoli , 2011) and a time evolution of network global characteristics was studied in (He and Deem , 2010). Topological and clustering properties of multiplex network of various commodities were discussed in (Barigozzi *et al.*, 2010), and an ecological ranking based on the nestedness of countries and products was presented in (Ermann and Shepelyansky , 2013a).

The money exchange between countries defines a directed network. Therefore Google matrix analysis can be introduced in a natural way. PageRank and CheiRank algorithms can be easily applied to this network with a straightforward correspondence with imports and exports. Two-dimensional ranking, introduced in Sec. IV, gives an illustrative representation of global importance of countries in the WTN. The important element of Google ranking of WTN is its democratic treatment of all world countries, independently of their richness, that follows the main principle of the United Nations (UN).

A. Democratic ranking of countries

The WTN is a directed network that can be constructed considering countries as nodes and money exchange as links. We follow the definition of the WTN of (Ermann and Shepelyansky , 2011b) where trade information comes from (UN COMTRADE, 2011). These data include all trades between countries for different products (using Standard International Trade Classification of goods, SITC1) from 1962 to 2009.

All useful information of the WTN is expressed via the *money matrix* M, which definition, in terms of its matrix elements M_{ij}, is defined as the money transfer (in *USD*) from country j to country i in a given year. This definition can be applied to a given specific product or to *all commodities*, which represent the sum over all products.

In contrast to the binary adjacency matrix A_{ij} of WWW (as the ones analyzed in SVIII and SX for example) M has weighted elements. This corresponds to a case when there are in principle multiple number of links from j to i and this number is proportional to *USD* amount transfer. Such a situation appears in Sec. VI for Ulam networks and Sec. VII for Linux PCN with a main difference that for the WTN case there is a very large variation of mass matrix elements M_{ij}, related to the fact that there is a very strong variation of richness of various countries.

Google matrices G and G^* are constructed according to the usual rules and relation (1) with M_{ij} and its transposed: $S_{ij} = M_{ij}/m_j$ and $S_{ij} = M_{ji}/m_j^*$ where $S_{ij} = 1/N$ and $S_{ij}^* = 1/N$, if for a given j all elements $M_{ij} = 0$ and $M_{ji} = 0$ respectively. Here $m_j = \sum_i M_{ij}$ and $m_j^* = \sum_i M_{ji}$ are the total export and import mass for country j. Thus the sum in each column of G or G^* is equal to unity. In this way Google matrices G and G^* of WTN allow to treat all countries on equal grounds independently of the fact if a given country is rich or poor. This kind of analysis treats in a democratic way all world countries in consonance with the standards of the UN.

The probability distributions of ordered PageRank $P(K)$ and CheiRank $P^*(K^*)$ depend on their indexes in a rather similar way with a power law decay given by β. For the fit of top 100 countries and *all commodities* the average exponent value is close to $\beta = 1$ corresponding to the Zipf law (Zipf, 1949).

The distribution of countries on PageRank-CheiRank plane for trade in *all commodities* in year 2008 is shown in panels (a) and (b) of Fig. 35 at $\alpha = 0.5$. Even if the Google matrix approach is based on a democratic ranking of international trade, being independent of total amount of export-import and PIB for a given country, the top ranks K and K^* belong to the group of industrially developed countries. This means that these countries have efficient trade networks with optimally distributed trade flows. Another striking feature of global distribution is that it is concentrated along the main diagonal $K = K^*$. This feature is not present in other networks

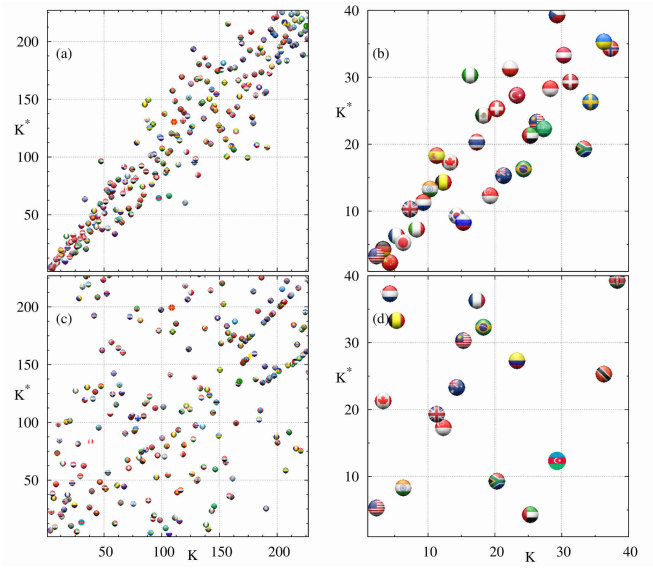

FIG. 35 (Color online) Country positions in PageRank-CheiRank plane (K, K^*) for world trade in various commodities in 2008. Each country is shown by circle with its own flag (for a better visibility the circle center is slightly displaced from its integer position (K, K^*) along direction angle $\pi/4$). The panels show the ranking for trade in the following commodities: *all commodities* (a) and (b); and *crude petroleum* (c) and (d). Panels (a) and (c) show a global scale with all 227 countries, while (b) and (d) give a zoom in the region of 40×40 top ranks. After (Ermann and Shepelyansky , 2011b).

studied before. The origin of this density concentration is related to a simple economy reason: for each country the total import is approximately equal to export since each country should keep in average an economic balance. This balance does not imply a symmetric money matrix, used in gravity model of trade (see e.g. (De Benedictis and Tajoli , 2011; Krugman *et al.*, 2011)), as can be seen in the significant broadening of distribution of Fig. 35 (especially at middle values of $K \sim 100$).

For a given country its trade is doing well if its $K^* < K$ so that the country exports more than it imports. The opposite relation $K^* > K$ corresponds to a bad trade situation (e.g. Greece being significantly above the diagonal). We also can say that local minima in the curve of $(K^* - K)$ *vs.* K correspond to a successful trade while maxima mark bad traders. In 2008 most successful were China, R of Korea, Russia, Singapore, Brazil, South Africa, Venezuela (in order of K for $K \leq 50$) while among bad traders we note UK, Spain, Nigeria, Poland, Czech Rep, Greece, Sudan with especially strong export drop for two last cases.

A comparison between local and global rankings of countries for both imports and exports gives a new tool to analyze countries economy. For example, in 2008 the most significant differences between CheiRank and the rank given by total exports are for *Canada* and *Mexico* with corresponding money export ranks $\tilde{K}^* = 11$ and 13 and with $K^* = 16$ and $K^* = 23$ respectively. These variations can be explained in the context that the export of these two countries is too strongly oriented on *USA*. In contrast *Singapore* moves up from $\tilde{K}^* = 15$ export position to $K^* = 11$ that shows the stability and broadness of its export trade, a similar situation appears for *India* moving up from $\tilde{K}^* = 19$ to $K^* = 12$ (see (Ermann and Shepelyansky , 2011b) for more detailed analysis).

B. Ranking of countries by trade in products

If we focus on the two-dimensional distribution of countries in a specific product we obtain a very different information. The symmetry approximately visible for *all commodities* is absolutely absent: the points are scattered practically over the whole square $N \times N$ (see Fig. 35). The reason of such a strong scattering is clear: e.g. for *crude petroleum* some countries export this product while other countries import it. Even if there is some flow from exporters to exporters it remains relatively low. This makes the Google matrix to be very asymmetric. Indeed, the asymmetry of trade flow is well visible in panels (c) and (d) of Fig. 35.

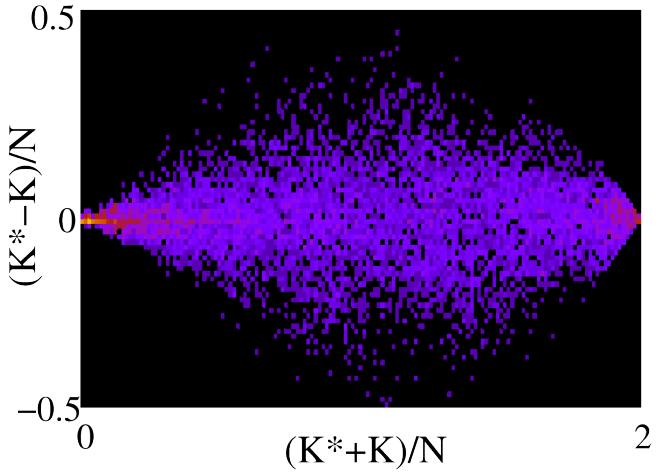

FIG. 36 (Color online) Spindle distribution for WTN of *all commodities* for all countries in the period 1962 - 2009 shown in the plane of $((K^* - K)/N, (K^* + K)/N)$ (coarse-graining inside each of 76×152 cells); data from the UN COMTRADE database. After (Ermann and Shepelyansky , 2011b).

The same comparison of global and local rankings done before for *all commodities* can be applied to specific products obtaining even more strong differences. For example for *crude petroleum* Russia moves up from $\tilde{K}^* = 2$ export position to $K^* = 1$ showing that its trade network in this product is better and broader than the one of Saudi Arabia which is at the first export position $\tilde{K}^* = 1$ in money volume. Iran moves in opposite direction from $\tilde{K}^* = 5$ money position down to $K^* = 14$ showing that its trade

network is restricted to a small number of nearby countries. A significant improvement of ranking takes place for Kazakhstan moving up from $\tilde{K}^* = 12$ to $K^* = 2$. The direct analysis shows that this happens due to an unusual fact that Kazakhstan is practically the only country which sells *crude petroleum* to the CheiRank leader in this product Russia. This puts Kazakhstan on the second position. It is clear that such direction of trade is more of political or geographical origin and is not based on economic reasons.

The same detailed analysis can be applied to all specific products given by SITC1. For example for trade of *cars* France goes up from $\tilde{K}^* = 7$ position in exports to $K^* = 3$ due to its broad export network.

C. Ranking time evolution and crises

The WTN has evolved during the period 1962 - 2009. The number of countries is increased by 38%, while the number of links per country for *all commodities* is increased in total by 140% with a significant increase from 50% to 140% during the period 1993 - 2009 corresponding to economy globalization. At the same time for a specific commodity the average number of links per country remains on a level of 3-5 links being by a factor 30 smaller compared to *all commodities* trade. During the whole period the total amount M_T of trade in *USD* shows an average exponential growth by 2 orders of magnitude.

A statistical density distribution of countries in the plane $(K^* - K, K^* + K)$ in the period 1962 - 2009 for *all commodities* is shown in Fig. 36. The distribution has a form of *spindle* with maximum density at the vertical axis $K^* - K = 0$. We remind that good exporters are on the lower side of this axis at $K^* - K < 0$, while the good importers (bad exporters) are on the upper side at $K^* - K > 0$.

The evolution of the ranking of countries for *all commodities* reflects their economical changes. The countries that occupy top positions tend to move very little in their ranks and can be associated to a *solid phase*. On the other hand, the countries in the middle region of $K^* + K$ have a gas like phase with strong rank fluctuations.

Examples of ranking evolution K and K^* for Japan, France, Fed R of Germany and Germany, Great Britain, USA, and for Argentina, India, China, USSR and Russian Fed are shown in Fig. 37. It is interesting to note that sharp increases in K mark crises in 1991, 1998 for Russia and in 2001 for Argentina (import is reduced in period of crises). It is also visible that in recent years the solid phase is perturbed by entrance of new countries like China and India. Other regional or global crisis could be highlighted due to the big fluctuations in the evolution of ranks. For example, in the range $81 \leq K + K^* \leq 120$, during the period of 1992 - 1998 some financial crises as Black Wednesday, Mexico crisis, Asian crisis and Russian crisis are appreciated with this ranking evolution.

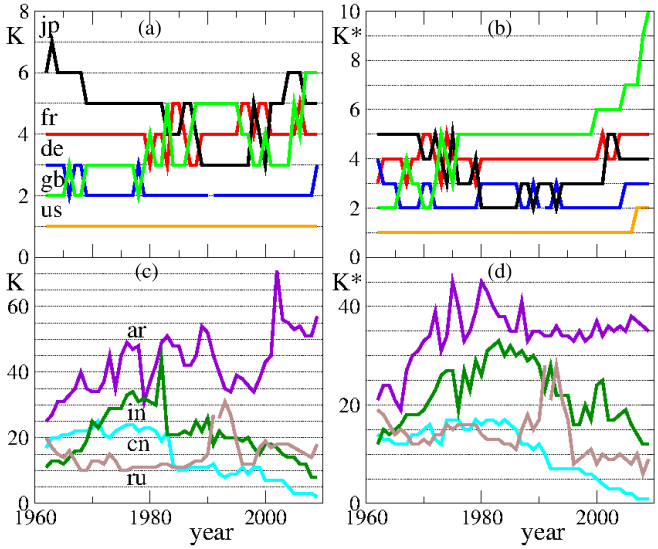

FIG. 37 (Color online) Time evolution of CheiRank and PageRank indexes K, K^* for some selected countries for *all commodities*. The countries shown panels (a) and (b) are: Japan (jp-black), France (fr-red), Fed R of Germany and Germany (de - both in blue), Great Britain (gb - green), USA (us - orange) [curves from top to bottom in 1962 in (a)]. The countries shown panels (c) and (d) are: Argentina (ar - violet), India (in - dark green), China (cn - cyan), USSR and Russian Fed (ru - both in gray) [curves from top to bottom in 1975 in (c)]. After (Ermann and Shepelyansky , 2011b).

D. Ecological ranking of world trade

Interesting parallels between multi-product world trade and interactions between species in ecological systems has been traced in (Ermann and Shepelyansky , 2013a). This approach is based on analysis of strength of transitions forming the Google matrix for the multi-product world trade network.

Ecological systems are characterized by high complexity and biodiversity (May, 2001) linked to nonlinear dynamics and chaos emerging in the process of their evolution (Lichtenberg and Lieberman, 1992). The interactions between species form a complex network whose properties can be analyzed by the modern methods of scale-free networks. The analysis of their properties uses a concept of mutualistic networks and provides a detailed understanding of their features being linked to a high nestedness of these networks (Bastolla *et al.*, 2009; Burgos *et al.*, 2007, 2008; Saverda *et al.*, 2011). Using the UN COMTRADE database we show that a similar ecological analysis gives a valuable description of the world trade: countries and trade products are analogous to plants and pollinators, and the whole trade network is characterized by a high nestedness typical for ecological networks.

An important feature of ecological networks is that they are highly structured, being very different from randomly interacting species (Bascompte *et al.*, 2003). Recently is has been shown that the mutualistic networks

between plants and their pollinators (Bascompte *et al.*, 2003; Memmott *et al.*, 2004; Olesen *et al.*, 2007; Rezende *et al.*, 2007; Vázquez and Aizen , 2004) are characterized by high nestedness which minimizes competition and increases biodiversity (Bastolla *et al.*, 2009; Burgos *et al.*, 2007, 2008; Saverda *et al.*, 2011).

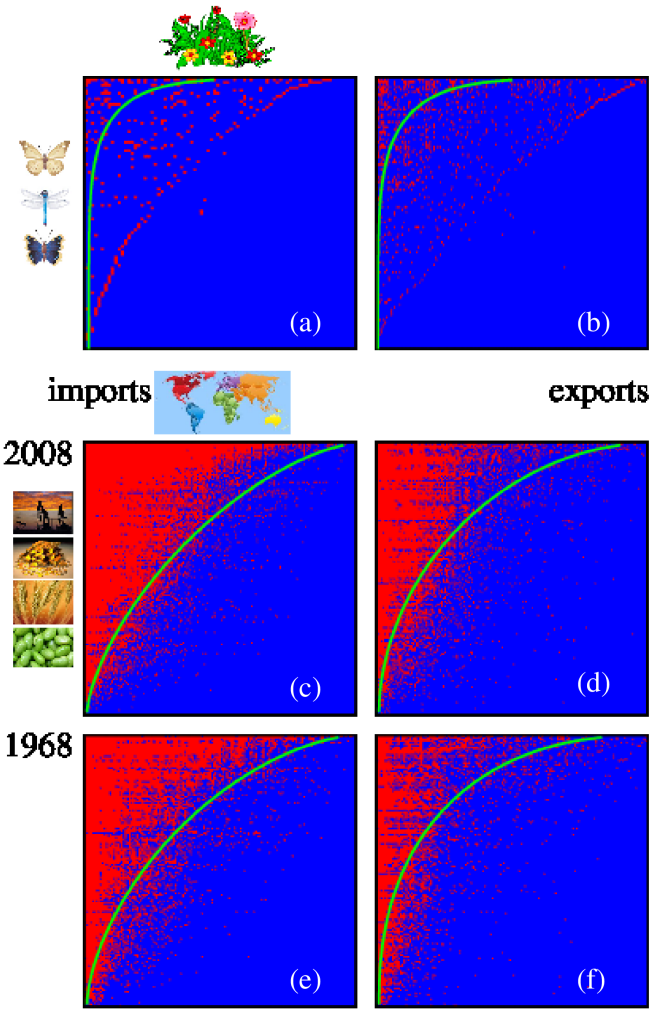

FIG. 38 (Color online) Nestedness matrices for the plant-animal mutualistic networks on top panels, and for the WTN of countries-products on middle and bottom panels. Panels (a) and (b) represent data of *ARR1* and *WES* networks from (Rezende *et al.*, 2007). The WTN matrices are computed with the threshold $\mu = 10^{-3}$ and corresponding $\varphi \approx 0.2$ for years 2008 (c,d) and 1968 (e,f) and 2008 for import (c,e) and export (d,f) panels. Red/gray and blue/black represent unit and zero elements respectively; only lines and columns with nonzero elements are shown. The order of plants-animals, countries-products is given by the nestedness algorithm (Rodríguez-Gironés *et al.*, 2006), the perfect nestedness is shown by green/gray curves for the corresponding values of φ. After (Ermann and Shepelyansky , 2013a).

The mutualistic WTN is constructed on the basis of the UN COMTRADE database from the matrix of trade transactions $M^p_{c',c}$ expressed in USD for a given product (commodity) p from country c to country c' in a given year (from 1962 to 2009). For product classification we use 3–digits SITC Rev.1 discussed above with the number of products $N_p = 182$. All these products are described in (UN COMTRADE, 2011) in the commodity code document SITC Rev1. The number of countries varies between $N_c = 164$ in 1962 and $N_c = 227$ in 2009. The import and export trade matrices are defined as $M^{(i)}_{p,c} = \sum_{c'=1}^{N_c} M^p_{c,c'}$ and $M^{(e)}_{p,c} = \sum_{c'=1}^{N_c} M^p_{c',c}$ respectively. We use the dimensionless matrix elements $m^{(i)} = M^{(i)}/M_{max}$ and $m^{(e)} = M^{(e)}/M_{max}$ where for a given year $M_{max} = max\{max[M^{(i)}_{p,c}], max[M^{(e)}_{p,c}]\}$. The distribution of matrix elements $m^{(i)}$, $m^{(e)}$ in the plane of indexes p and c, ordered by the total amount of import/export in a decreasing order, are shown and discussed in (Ermann and Shepelyansky , 2013a). In global, the distributions of $m^{(i)}$, $m^{(e)}$ remain stable in time especially in a view of 100 times growth of the total trade volume during the period 1962-2009. The fluctuations of $m^{(e)}$ are larger compared to $m^{(i)}$ case since certain products, e.g. petroleum, are exported by only a few countries while it is imported by almost all countries.

To use the methods of ecological analysis we construct the mutualistic network matrix for import $Q^{(i)}$ and export $Q^{(e)}$ whose matrix elements take binary value 1 or 0 if corresponding elements $m^{(i)}$ and $m^{(e)}$ are respectively larger or smaller than a certain trade threshold value μ. The fraction φ of nonzero matrix elements varies smoothly in the range $10^{-6} \leq \mu \leq 10^{-2}$ and the further analysis is not really sensitive to the actual μ value inside this broad range.

In contrast to ecological systems (Bastolla *et al.*, 2009) the world trade is described by a directed network and hence we characterize the system by two mutualistic matrices $Q^{(i)}$ and $Q^{(e)}$ corresponding to import and export. Using the standard nestedness BINMATNEST algorithm (Rodríguez-Gironés *et al.*, 2006) we determine the nestedness parameter η of the WTN and the related nestedness temperature $T = 100(1 - \eta)$. The algorithm reorders lines and columns of a mutualistic matrix concentrating nonzero elements as much as possible in the top left corner and thus providing information about the role of immigration and extinction in an ecological system. A high level of nestedness and ordering can be reached only for systems with low T. It is argued that the nested architecture of real mutualistic networks increases their biodiversity.

The nestedness matrices generated by the BINMATNEST algorithm (Rodríguez-Gironés *et al.*, 2006) are shown in Fig. 38 for ecology networks ARR1 ($N_{pl} = 84$, $N_{anim} = 101$, $\varphi = 0.043$, $T = 2.4$) and WES ($N_{pl} = 207$, $N_{anim} = 110$, $\varphi = 0.049$, $T = 3.2$) from (Rezende *et al.*, 2007). Using the same algorithm we generate the nestedness matrices of WTN using the mutualistic matrices for import $Q^{(i)}$ and export $Q^{(i)}$ for the WTN in years 1968 and 2008 using a fixed typical threshold $\mu = 10^{-3}$ (see Fig. 38). As for ecological systems, for the WTN data we also obtain rather small nestedness

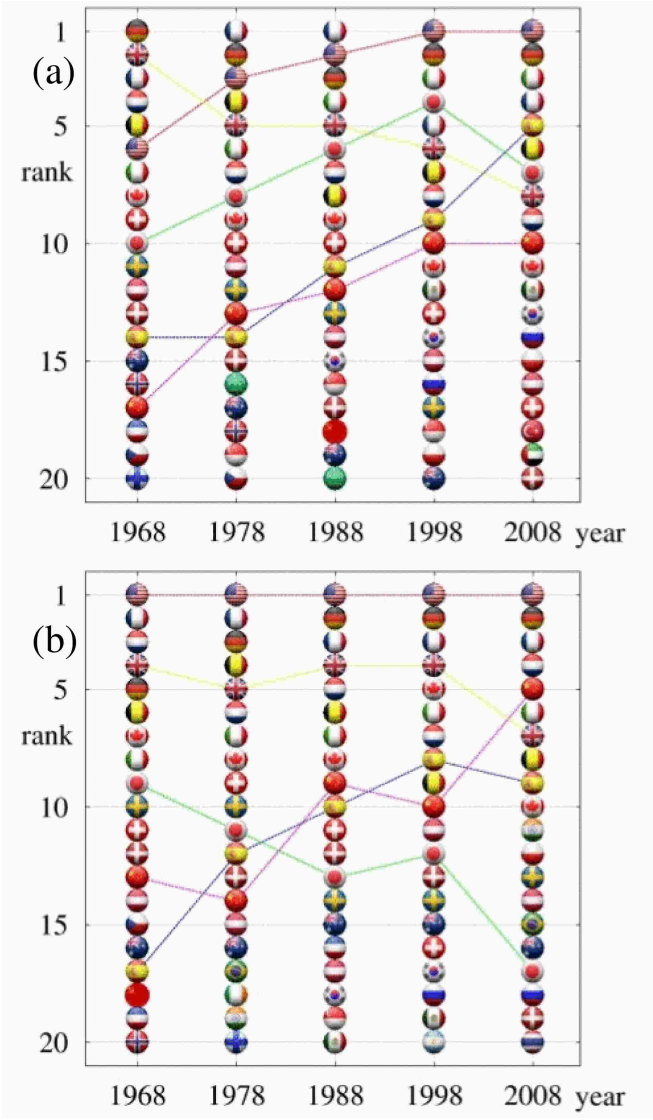

FIG. 39 (Color online) Top 20 EcoloRank countries as a function of years for the WTN import (a) and export (b) panels. The ranking is given by the nestedness algorithm for the trade threshold $\mu = 10^{-3}$; each country is represented by its corresponding flag. As an example, dashed lines show time evolution of the following countries: USA, UK, Japan, China, Spain. After (Ermann and Shepelyansky , 2013a).

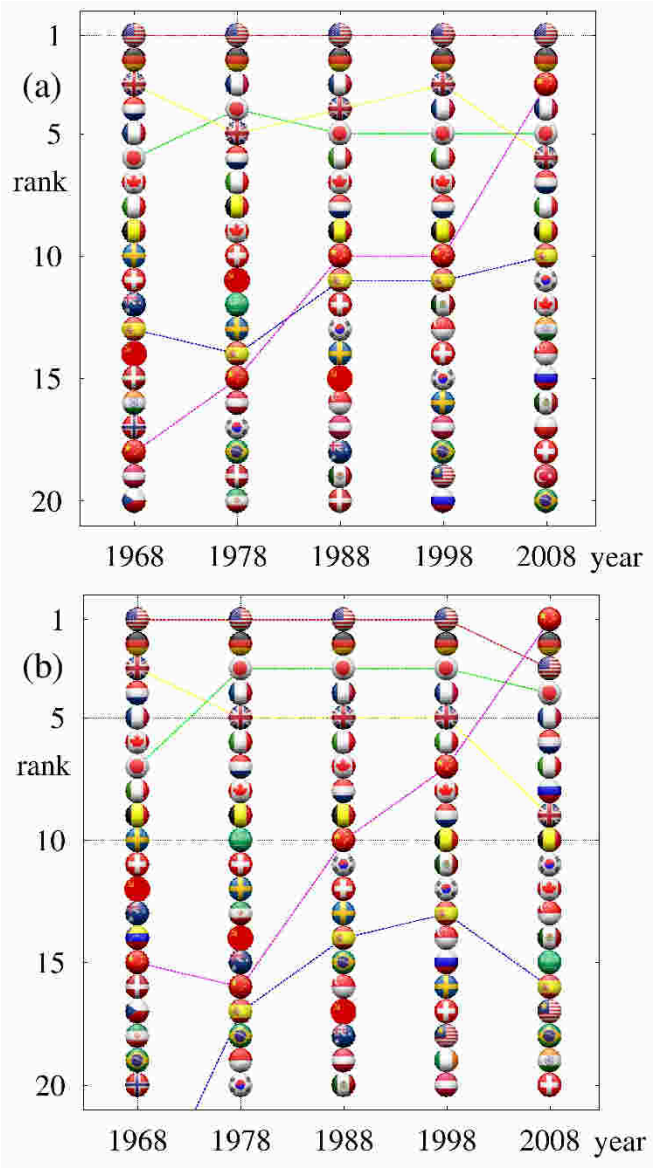

FIG. 40 (Color online) Top 20 countries as a function of years ranked by the total monetary trade volume of the WTN in import (a) and export (b) panels respectively; each country is represented by its corresponding flag. Dashed lines show time evolution of the same countries as in Fig. 39. After (Ermann and Shepelyansky , 2013a).

temperature ($T \approx 6/8$ for import/export in 1968 and $T \approx 4/8$ in 2008 respectively). These values are by a factor 9/4 of times smaller than the corresponding T values for import/export from random generated networks with the corresponding values of φ.

The small value of nestedness temperature obtained for the WTN confirms the validity of the ecological analysis of WTN structure: trade products play the role of pollinators which produce exchange between world countries, which play the role of plants. Like in ecology the WTN evolves to the state with very low nestedness temperature that satisfies the ecological concept of system stability appearing as a result of high network nestedness (Bastolla *et al.*, 2009).

The nestedness algorithm creates effective ecological ranking (EcoloRanking) of all UN countries. The evolution of 20 top ranks throughout the years is shown in Fig. 39 for import and export. This ranking is quite different from the more commonly applied ranking of countries by their total import/export monetary trade volume (CIA, 2009) (see corresponding data in Fig. 40) or the democratic ranking of WTN based on the Google matrix

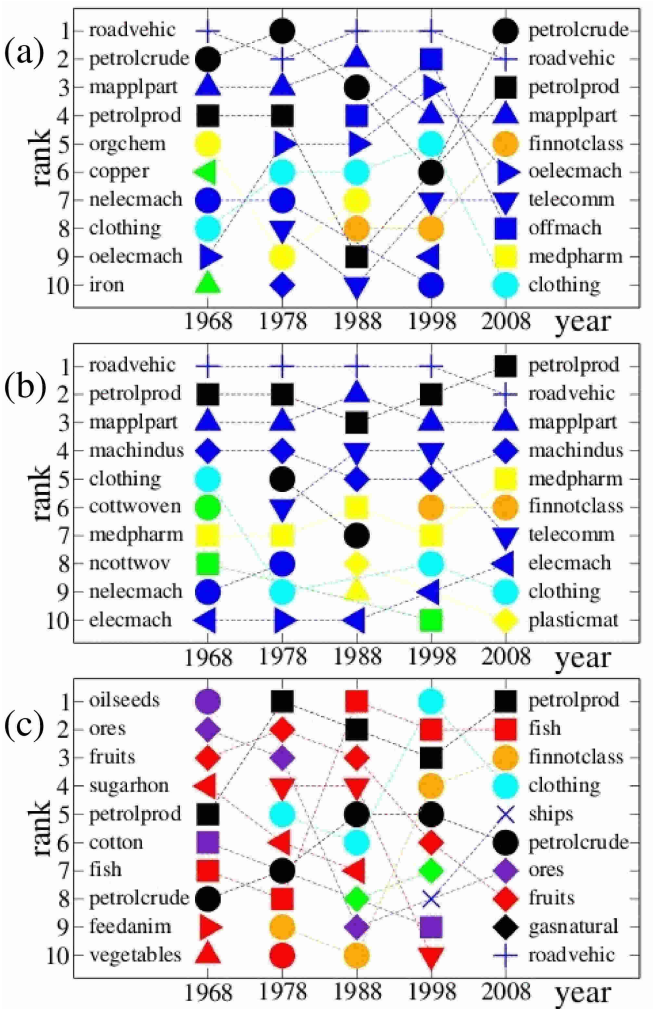

FIG. 41 (Color online) Top 10 ranks of trade products as a function of years for the WTN. Panel (a): ranking of products by monetary trade volume. Panels (b), (c): ranking is given by the nestedness algorithm for import (b) and export (c) with the trade threshold $\mu = 10^{-3}$. Each product is shown by its own symbol with short name written at years 1968, 2008; symbol color marks 1st SITC digit; SITC codes of products and their names are given in (UN COMTRADE, 2011) and Table 2 in (Ermann and Shepelyansky , 2013a). After (Ermann and Shepelyansky , 2013a).

analysis discussed above. Indeed, in 2008 China is at the top rank for total export volume but it is only at 5th position in EcoloRank (see Fig. 39, Fig. 40). In a similar way Japan moves down from 4th to 17th position while USA raises up from 3rd to 1st rank.

The same nestedness algorithm generates not only the ranking of countries but also the ranking of trade products for import and export which is presented in Fig. 41. For comparison we also show there the standard ranking of products by their trade volume. In Fig. 41 the color of symbol marks the 1st SITC digit described in figure, (UN COMTRADE, 2011) and Table 2 in (Ermann and Shepelyansky , 2013a).

The origin of such a difference between EcoloRanking and trade volume ranking of countries is related to the main idea of mutualistic ranking in ecological systems: the nestedness ordering stresses the importance of mutualistic pollinators (products for WTN) which generate links and exchange between plants (countries for WTN). In this way generic products, which participate in the trade between many countries, become of primary importance even if their trade volume is not at the top lines of import or export. In fact such mutualistic products glue the skeleton of the world trade while the nestedness concept allows to rank them in order of their importance. The time evolution of this EcoloRanking of products of WTN is shown in Fig. 41 for import/export in comparison with the product ranking by the monetary trade volume (since the trade matrix is diagonal in product index the ranking of products in the latter case is the same for import/export). The top and middle panels have dominate colors corresponding to machinery (SITC Rev. 1 code 7; blue) and mineral fuels (3; black) with a moderate contribution of chemicals (5; yellow) and manufactured articles (8; cyan) and a small fraction of goods classified by material (6; green). Even if the global structure of product ranking by trade volume has certain similarities with import EcoloRanking there are also important new elements. Indeed, in 2008 the mutualistic significance of petroleum products (code 332), *machindus* (machines for special industries code 718) and *medpharm* (medical-pharmaceutic products code 541) is much higher compared to their volume ranking, while petroleum crude (code 331) and office machines (code 714) have smaller mutualistic significance compared to their volume ranking.

The new element of EcoloRanking is that it differentiates between import and export products while for trade volume they are ranked in the same way. Indeed, the dominant colors for export (Fig. 41 bottom panel) correspond to food (SITC Rev. 1 code 0; red) with contribution of black (present in import) and crude materials (code 2; violet); followed by cyan (present in import) and more pronounced presence of *finnotclass* (commodities/transactions not classified code 9; brown). EcoloRanking of export shows a clear decrease tendency of dominance of SITC codes 0 and 2 with time and increase of importance of codes 3,7. It is interesting to note that the code 332 of petroleum products is vary vulnerable in volume ranking due to significant variations of petroleum prices but in EcoloRanking this product keeps the stable top positions in all years showing its mutualistic structural importance for the world trade. EcoloRanking of export shows also importance of fish (code 031), clothing (code 841) and fruits (code 051) which are placed on higher positions compared to their volume ranking. At the same time *roadvehic* (code 732), which are at top volume ranking, have relatively low ranking in export since only a few countries dominate the production of road vehicles.

It is interesting to note that in Fig. 41 petroleum crude is at the top of trade volume ranking e.g. in 2008 (top panel) but it is absent in import EcoloRanking (middle panel) and it is only on 6th position in export EcoloRanking (bottom panel). A similar feature is visible for years 1968, 1978. On a first glance this looks surprising but in fact for mutualistic EcoloRanking it is important that a given product is imported from top EcoloRank countries: this is definitely not the case for petroleum crude which practically is not produced inside top 10 import EcoloRank countries (the only exception is USA, which however also does not export much). Due to that reason this product has low mutualistic significance.

The mutualistic concept of product importance is at the origin of significant difference of EcoloRanking of countries compared to the usual trade volume ranking (see Fig. 39, Fig. 40). Indeed, in the latter case China and Japan are at the dominant positions but their trade is concentrated in specific products which mutualistic role is relatively low. In contrast USA, Germany and France keep top three EcoloRank positions during almost 40 years clearly demonstrating their mutualistic power and importance for the world trade.

Thus our results show the universal features of ecologic ranking of complex networks with promising future applications to trade, finance and other areas.

E. Remarks on world trade and banking networks

The new approach to the world trade, based on the Google matrix analysis, gives a democratic type of ranking being independent of the trade amount of a given country. In this way rich and poor countries are treated on equal democratic grounds. In a certain sense PageRank probability for a given country is proportional to its rescaled import flows while CheiRank is proportional to its rescaled export flows inside of the WTN.

The global characteristics of the world trade are analyzed on the basis of this new type of ranking. Even if all countries are treated now on equal democratic grounds still we find at the top rank the group of industrially developed countries approximately corresponding to *G-20* and recover 74% of countries listed in *G-20*. The Google matrix analysis demonstrates an existence of two solid state domains of rich and poor countries which remain stable during the years of consideration. Other countries correspond to a gas phase with ranking strongly fluctuating in time. We propose a simple random matrix model which well describes the statistical properties of rank distribution for the WTN (Ermann and Shepelyansky , 2011b).

The comparison between usual ImportRank–ExportRank (see e.g. (CIA, 2009)) and our PageRank–CheiRank approach shows that the later highlights the trade flows in a new useful manner which is complementary to the usual analysis. The important difference between these two approaches is due to the fact that ImportRank–ExportRank method takes into account only global amount of money exchange between a country and the rest of the world while PageRank–CheiRank approach takes into account all links and money flows between all countries.

The future developments should consider a matrix with all countries and all products which size becomes significantly larger ($N \sim 220 \times 10^4 \sim 2 \times 10^6$) comparing to a modest size $N \approx 227$ considered here. However, some new problems of this multiplex network analysis should be resolved combining a democracy in countries with volume importance of products which role is not democratic. It is quite possible that such an improved analysis will generate an asymmetric ranking of products in contrast to their symmetric ranking by volume in export and import. The ecological ranking of the WTN discussed in the previous SubSec. indicates preferences and asymmetry of trade in multiple products (Ermann and Shepelyansky , 2013a).

It is also important to note that usually in economy researchers analyze time evolution of various indexes studying their correlations. The results presented above for the WTN show that in addition to time evolution there is also evolution in space of the network. Like for waves in an ocean time and space are both important and we think that time and space study of trade captures important geographical factors which will play a dominant role for analysis of contamination propagation over the WTN in case of crisis. We think that the WTN data capture many essential elements which will play a rather similar role for financial flows in the interbank payment networks. We expect that the analysis of financial flows between bank units would prevent important financial crisis shaking the world in last years. Unfortunately, in contrast to WWW and UN COMTRADE, the banks keep hidden their financial flows. Due to this secrecy of banks the society is still suffering from financial crises. And all this for a network of very small size estimated on a level of 50 thousands bank units for the whole world being by a factor million smaller than the present size of WWW (e.g. Fedwire interbank payment network of USA contains only 6600 nodes (Soramaki *et al.*, 2007)). In a drastic contrast with bank networks the WWW provided a public access to its nodes changing the world on a scale of 20 years. A creation of the World Bank Web (WBW) with information accessible for authorized investigators would allow to understand and control financial flows in an efficient manner preventing the society from bank crises. We note that the methods of network analysis and ranking start to attract interest of researchers in various banks (see e.g. (Craig and von Peter, 2010; Garratt *et al.*, 2011)).

XII. NETWORKS WITH NILPOTENT ADJACENCY MATRIX

A. General properties

In certain networks (Frahm *et al.*, 2012a, 2014b) it is possible to identify an ordering scheme for the nodes such that the adjacency matrix has non-vanishing elements A_{mn} only for nodes $m < n$ providing a triangular matrix structure. In these cases it is possible to provide a semi-analytical theory (Frahm *et al.*, 2012a, 2014b) which allows to simplify the numerical calculation of the non-vanishing eigenvalues of the matrix S introduced in Sec. III.A. It is useful to write this matrix in the form

$$S = S_0 + (1/N)\, e\, d^T \tag{10}$$

where the vector e has unit entries for all nodes and the *dangling vector* d has unit entries for dangling nodes and zero entries for the other nodes. The extra contribution $e\, d^T/N$ just replaces the empty columns (of S_0) with $1/N$ entries at each element. For a triangular network structure the matrix S_0 is nilpotent, i.e. $S_0^l = 0$ for some integer $l > 0$ and $S_0^{l-1} \neq 0$. Furthermore for the network examples studied previously (Frahm *et al.*, 2012a, 2014b) we have $l \ll N$ which has important consequences for the eigenvalue spectrum of S.

There are two groups of (right) eigenvectors ψ of S with eigenvalue λ. For the first group the quantity $C = d^T \psi$ vanishes and ψ is also an eigenvector of S_0 and if S_0 is nilpotent we have $\lambda = 0$ (there are also many higher order generalized eigenvectors associated to $\lambda = 0$). For the second group we have $C \neq 0$, $\lambda \neq 0$ and the eigenvector is given by $\psi = (\lambda \mathbb{1} - S_0)^{-1}\, C\, e/N$. Expanding the matrix inverse in a finite geometric series (for nilpotent S_0) and applying the condition $C = d^T \psi$ on this expression one finds that the eigenvalue must be a zero of the *reduced polynomial* of degree l:

$$\mathcal{P}_r(\lambda) = \lambda^l - \sum_{j=0}^{l-1} \lambda^{l-1-j}\, c_j = 0\ , \quad c_j = d^T\, S_0^j\, e/N\ . \tag{11}$$

This shows that there are at most l non-vanishing eigenvalues of S with eigenvectors $\psi \propto \sum_{j=0}^{l-1} \lambda^{-j-1}\, v^{(j)}$ where $v^{(j)} = S_0^j\, e/N$ for $j = 0, \ldots, l-1$. Actually, the vectors $v^{(j)}$ generate an S-invariant l-dimensional subspace and from $S\, v^{(j)} = c_j\, v^{(0)} + v^{(j+1)}$ (using the identification $v^{(l)} = 0$) one obtains directly the $l \times l$ representation matrix $\bar{S}$ of S with respect to $v^{(j)}$ (Frahm *et al.*, 2012a). Furthermore, the characteristic polynomial of $\bar{S}$ is indeed given by the reduced polynomial (11) and the sum rule $\sum_{j=0}^{l-1} c_j = 1$ ensures that $\lambda = 1$ is indeed a zero of $\mathcal{P}_r(\lambda)$ (Frahm *et al.*, 2012a). The corresponding eigenvector (PageRank P at $\alpha = 1$) is given by $P \propto \sum_{j=0}^{l-1} v^{(j)}$. The remaining $N - l$ (generalized) eigenvectors of S are associated to many different Jordan blocks of S_0 for the eigenvalue $\lambda = 0$.

These l non-vanishing complex eigenvalues can be numerically computed as the zeros of the reduced polynomial by the Newton-Maehly method, by a numerical diagonalization of the "small" representation matrix $\bar{S}$ (or better a more stable transformed matrix with identical eigenvalues) or by the Arnoldi method using the uniform vector e as initial vector. In the latter case the Arnoldi method should theoretically (in absence of rounding errors) exactly explore the l-dimensional subspace of the vectors $v^{(j)}$ and break off after l iterations with l exact eigenvalues.

However, numerical rounding errors may have a strong effect due to the Jordan blocks for the zero eigenvalue (Frahm *et al.*, 2012a). Indeed, an error ϵ appearing in a left bottom corner of a Jordan matrix of size D with zero eigenvalue leads to numerically induced eigenvalues on a complex circle of radius

$$|\lambda_\epsilon| = \epsilon^{1/D}\ . \tag{12}$$

Such an error can become significant with $|\lambda| > 0.1$ even for $\epsilon \sim 10^{-15}$ as soon as $D > 15$. We call this phenomenon the Jordan error enhancement. Furthermore, also the numerical determination of the zeros of $\mathcal{P}_r(\lambda)$ for large values of $l \sim 10^2$ can be numerically rather difficult. Thus, it may be necessary to use a high precision library such as the GNU GMP library either for the determination of the zeros of $\mathcal{P}_r(\lambda)$ or for the Arnoldi method (Frahm *et al.*, 2014b).

B. PageRank of integers

A network for integer numbers (Frahm *et al.*, 2012a) can be constructed by linking an integer number $n \in \{1, \ldots, N\}$ to its divisors m different from 1 and n itself by an adjacency matrix $A_{mn} = M(n, m)$ where the multiplicity $M(n, m)$ is the number of times we can divide n by m, i.e. the largest integer such that $m^{M(n,m)}$ is a divisor of n, and $A_{mn} = 0$ for all other cases. The number 1 and the prime numbers are not linked to any other number and correspond to dangling nodes. The total size N of the matrix is fixed by the maximal considered integer. According to numerical data the number of links $N_\ell = \sum_{mn} A_{mn}$ is approximately given by $N_\ell = N\,(a_\ell + b_\ell \ln N)$ with $a_\ell = -0.901 \pm 0.018$, $b_\ell = 1.003 \pm 0.001$.

The matrix elements A_{mn} are different from zero only for $n \geq 2m$ and the associated matrix S_0 is therefore nilpotent with $S_0^l = 0$ and $l = [\log_2(N)] \ll N$. This triangular matrix structure can be seen in Fig. 42(a) which shows the amplitudes of S. The vertical green/gray lines correspond to the extra contribution due to the dangling nodes. These l non-vanishing eigenvalues of S can be efficiently calculated as the zeros of the reduced polynomial (11) up to $N = 10^9$ with $l = 29$. For $N = 10^9$ the largest eigenvalues are $\lambda_1 = 1$, $\lambda_{2,3} \approx -0.27178 \pm i\, 0.42736$, $\lambda_4 \approx -0.17734$ and $|\lambda_j| < 0.1$ for $j \geq 5$. The dependence of the eigenvalues on N seems to scale with

the parameter $1/\ln(N)$ for $N \to \infty$ and in particular $\gamma_2(N) = -2\ln|\lambda_2(N)| \approx 1.020 + 7.14/\ln N$ (Frahm *et al.*, 2012a). Therefore the first eigenvalue is clearly separated from the second eigenvalue and one can chose the damping factor $\alpha = 1$ without any problems to define a unique PageRank.

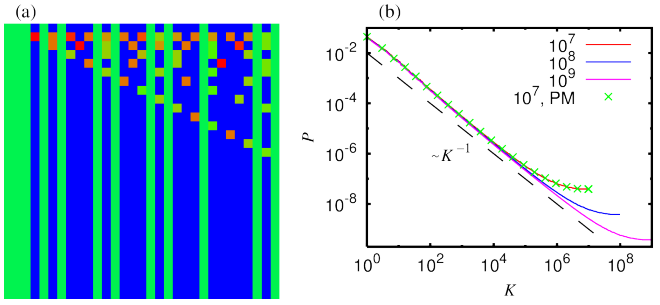

FIG. 42 (Color online) Panel (a): the Google matrix of integers, the amplitudes of matrix elements S_{mn} are shown by color with blue/black for minimal zero elements and red/gray for maximal unity elements, with $1 \le n \le 31$ corresponding to $x-$axis (with $n = 1$ corresponding to the left column) and $1 \le m \le 31$ for $y-$axis (with $m = 1$ corresponding to the upper row). Panel (b): the full lines correspond to the dependence of PageRank probability $P(K)$ on index K for the matrix sizes $N = 10^7$, 10^8, 10^9 with the PageRank evaluated by the exact expression $P \propto \sum_{j=0}^{l-1} v^{(j)}$. The green/gray crosses correspond to the PageRank obtained by the power method for $N = 10^7$; the dashed straight line shows the Zipf law dependence $P \sim 1/K$. After (Frahm *et al.*, 2012a).

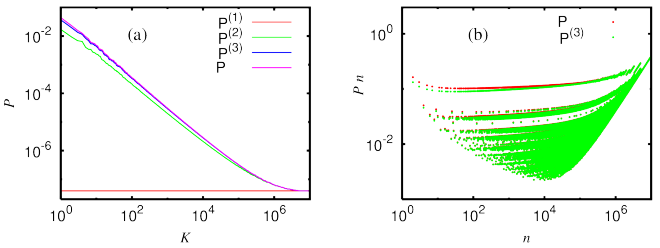

FIG. 43 (Color online) Panel (a): comparison of the first three PageRank approximations $P^{(i)} \propto \sum_{j=0}^{i-1} v^{(j)}$ for $i = 1, 2, 3$ and the exact PageRank dependence $P(K)$. Panel (b): comparison of the dependence of the rescaled probabilities nP and $nP^{(3)}$ on n. Both panels correspond to the case $N = 10^7$. After (Frahm *et al.*, 2012a).

The large values of N are possible because the vector iteration $v^{(j+1)} = S_0\, v^{(j)}$ can actually be computed without storing the $N_\ell \sim N \ln N$ non-vanishing elements of S_0 by using the relation:

$$v_n^{(j+1)} = \sum_{m=2}^{[N/n]} \frac{M(mn,m)}{Q(mn)}\, v_{mn}^{(j)}\,, \quad \text{if } n \ge 2 \qquad (13)$$

and $v_1^{(j+1)} = 0$ (Frahm *et al.*, 2012a). The initial vector is given by $v^{(0)} = e/N$ and $Q(n) = \sum_{m=2}^{n-1} M(n,m)$ is the number of divisors of n (taking into account the multiplicity). The multiplicity $M(mn,n)$ can be recalculated during each iteration and one needs only to store $N(\ll N_\ell)$ integer numbers $Q(n)$. It is also possible to reformulate (13) in a different way without using $M(mn,n)$ (Frahm *et al.*, 2012a). The vectors $v^{(j)}$ allow to compute the coefficients $c_j = d^T v^{(j)}$ in the reduced polynomial and the PageRank $P \propto \sum_{j=0}^{l-1} v^{(j)}$. Fig. 42(b) shows the PageRank for $N \in \{10^7, 10^8, 10^9\}$ obtained in this way and for comparison also the result of the power method for $N = 10^7$.

Actually Fig. 43 shows that in the sum $P \propto \sum_{j=0}^{l-1} v^{(j)}$ already the first three terms give a quite satisfactory approximation to the PageRank allowing a further analytical simplified evaluation (Frahm *et al.*, 2012a) with the result $P(n) \approx C_N/(b_n\, n)$ for $n \ll N$, where C_N is the normalization constant and $b_n = 2$ for prime numbers n and $b_n = 6 - \delta_{p_1,p_2}$ for numbers $n = p_1\, p_2$ being a product of two prime numbers p_1 and p_2. The behavior $P(n)\, n \approx C_N/b_n$, which takes approximately constant values on several branches, is also visible in Fig. 43 with C_N/b_n decreasing if n is a product of many prime numbers. The numerical results up to $N = 10^9$ show that the numbers n, corresponding to the leading PageRank values for $K = 1, 2, \ldots, 32$, are $n = 2, 3, 5, 7, 4, 11, 13, 17, 6, 19, 9, 23, 29, 8, 31, 10, 37, 41, 43, 14, 47, 15, 53, 59, 61, 25, 67, 12, 71, 73, 22, 21$ with about 30% of non-primes among these values (Frahm *et al.*, 2012a).

A simplified model for the network for integer numbers with $M(n,m) = 1$ if m is divisor of n and $1 < m < n$ has also been studied with similar results (Frahm *et al.*, 2012a).

C. Citation network of Physical Review

Citation networks for Physical Review and other scientific journals can be defined by taking published articles as nodes and linking an article A to another article B if A cites B. PageRank and similar analysis of such networks are efficient to determine influential articles (Newman , 2001; Radicchi *et al.*, 2009; Redner , 1998, 2005).

In citation network links go mostly from newer to older articles and therefore such networks have, apart from the dangling node contributions, typically also a (nearly) triangular structure as can be seen in Fig. 44 which shows a coarse-grained density of the corresponding Google matrix for the citation network of Physical Review from the very beginning until 2009 (Frahm *et al.*, 2014b). However, due to the delay of the publication process in certain rare instances a published paper may cite another paper that is actually published a little later and sometimes two papers may even cite mutually each other. Therefore the matrix structure is not exactly triangular but in the coarse-grained density in Fig. 44 the rare "future citations" are not well visible.

The nearly triangular matrix structure implies large

dimensional Jordan blocks associated to the eigenvalue $\lambda = 0$. This creates the Jordan error enhancement (12) with severe numerical problems for accurate computation of eigenvalues in the range $|\lambda| < 0.3 - 0.4$ when using the Arnoldi method with standard double-precision arithmetic (Frahm *et al.*, 2014b).

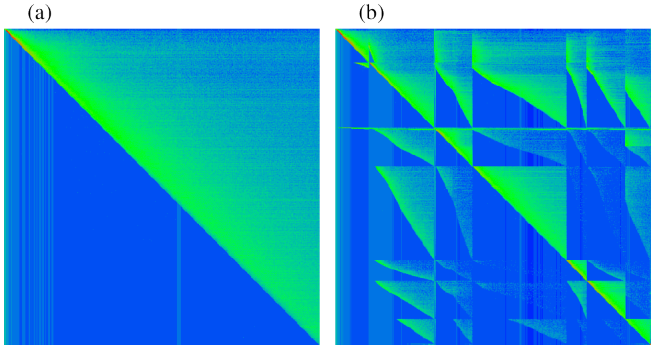

FIG. 44 (Color online) Different representations of the Google matrix structure for the Physical Review network until 2009. (a) Density of matrix elements $G_{tt'}$ in the basis of the publication time index t (and t'). (b) Density of matrix elements in the basis of journal ordering according to: Phys. Rev. Series I, Phys. Rev., Phys. Rev. Lett., Rev. Mod. Phys., Phys. Rev. A, B, C, D, E, Phys. Rev. STAB and Phys. Rev. STPER. and with time index ordering inside each journal. Note that the journals Phys. Rev. Series I, Phys. Rev. STAB and Phys. Rev. STPER are not clearly visible due to a small number of published papers. Also Rev. Mod. Phys. appears only as a thick line with 2-3 pixels (out of 500) due to a limited number of published papers. The different blocks with triangular structure correspond to clearly visible seven journals with considerable numbers of published papers. Both panels show the coarse-grained density of matrix elements on 500×500 square cells for the entire network. Color shows the density of matrix elements (of G at $\alpha = 1$) changing from blue/black for minimum zero value to red/gray at maximum value. After (Frahm *et al.*, 2014b).

One can eliminate the small number of future citations (12126 which is 0.26 % of the total number of links $N_\ell = 4691015$) and determine the complex eigenvalue spectrum of a triangular reduced citation network using the semi-analytical theory presented in previous subsection. It turns out that in this case the matrix S_0 is nilpotent $S_0^l = 0$ with $l = 352$ which is much smaller than the total network size $N = 463348$. The 352 non-vanishing eigenvalues can be determined numerically as the zeros of the polynomial (11) but due to an alternate sign problem with a strong loss of significance it is necessary to use the high precision library GMP with 256 binary digits (Frahm *et al.*, 2014b).

The semi-analytical theory can also be generalized to the case of *nearly* triangular networks, i.e. the full citation network including the future citations. In this case the matrix S_0 is no longer nilpotent but one can still generalize the arguments of previous subsection and discuss the two cases where the quantity $C = d^T \psi$ either vanishes (eigenvectors of first group) or is different from zero (eigenvectors of second group). The eigenvalues λ for the first group, which may now be different from zero, can be determined by a quite complicated but numerically very efficient procedure using the subspace eigenvalues of S and degenerate subspace eigenvalues of S_0 (due to absence of dangling node contributions the matrix S_0 produces much larger invariant subspaces than S) (Frahm *et al.*, 2014b). The eigenvalues of the second group are given as the complex zeros of the rational function:

$$\mathcal{R}(\lambda) = 1 - d^T \frac{\mathbb{1}}{\lambda \mathbb{1} - S_0} e/N = 1 - \sum_{j=0}^{\infty} c_j \lambda^{-1-j} \qquad (14)$$

with c_j given as in (11) and now the series is not finite since S_0 is not nilpotent. For the citation network of Physical Review the coefficients c_j behave as $c_j \propto \rho_1^j$ where $\rho_1 \approx 0.902$ is the largest eigenvalue of the matrix S_0 with an eigenvector non-orthogonal to d. Therefore the series in (14) converges well for $|\lambda| > \rho_1$ but in order to determine the spectrum the rational function $\mathcal{R}(\lambda)$ needs to be evaluated for smaller values of $|\lambda|$. This problem can be solved by interpolating $\mathcal{R}(\lambda)$ with (another) rational function using a certain number of support points on the complex unit circle, where (14) converges very well, and determining the complex zeros, well inside the unit circle, of the numerator polynomial using again the high precision library GMP (Frahm *et al.*, 2014b). In this way using 16384 binary digits one may obtain 2500 reliable eigenvalues of the second group.

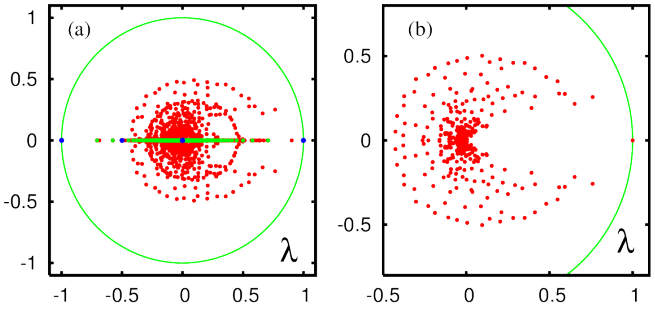

FIG. 45 (Color online) (a) Most accurate spectrum of eigenvalues for the full Physical Review network; red/gray dots represent the core space eigenvalues obtained by the rational interpolation method with the numerical precision of $p = 16384$ binary digits, $n_R = 2500$ eigenvalues; green (light gray) dots show the degenerate subspace eigenvalues of the matrix S_0 which are also eigenvalues of S with a degeneracy reduced by one (eigenvalues of the first group); blue/black dots show the direct subspace eigenvalues of S. (b) Spectrum of numerically accurate 352 non-vanishing eigenvalues of the Google matrix for the triangular reduced Physical Review network determined by the Newton-Maehly method applied to the reduced polynomial (11) with a high-precision calculation of 256 binary digits; note the absence of subspace eigenvalues for this case. In both panels the green/gray curve represents the unit circle. After (Frahm *et al.*, 2014b).

The numerical high precision spectra obtained by the semi-analytic methods for both cases, triangular reduced and full citation network, are shown in Fig. 45. One may mention that it is also possible to implement the Arnoldi method using the high precision library GMP for both cases and the resulting eigenvalues coincide very accurately with the semi-analytic spectra for both cases (Frahm *et al.*, 2014b).

When the spectrum of G is determined with a good accuracy we can test the validity of the fractal Weyl law (5) changing the matrix size N_t by considering articles published from the beginning to a certain time moment t measured in years. The data presented in Fig. 46 show that the network size grows approximately exponentially as $N_t = 2^{(t-t_0)/\tau}$ with the fit parameters $t_0 = 1791$, $\tau = 11.4$. The time interval considered in Fig. 46 is $1913 \leq t \leq 2009$ since the first data point corresponds to $t = 1913$ with $N_t = 1500$ papers published between 1893 and 1913. The results, for the number N_λ of eigenvalues with $|\lambda_i| > \lambda$, show that its growth is well described by the relation $N_\lambda = a\,(N_t)^\nu$ for the range when the number of articles becomes sufficiently large $3 \times 10^4 \leq N_t < 5 \times 10^5$. This range is not very large and probably due to that there is a certain dependence of the exponent ν on the range parameter λ_c. At the same time we note that the maximal matrix size N studied here is probably the largest one used in numerical studies of the fractal Weyl law. We have $0.47 < \nu < 0.6$ for all $\lambda_c \geq 0.4$ that is definitely smaller than unity and thus the fractal Weyl law is well applicable to the Phys. Rev. network. The value of ν increases up to 0.7 for the data points with $\lambda_c < 0.4$ but this is due to the fact here N_λ also includes some numerically incorrect eigenvalues related to the numerical instability of the Arnoldi method at standard double-precision (52 binary digits) as discussed above.

We conclude that the most appropriate choice for the description of the data is obtained at $\lambda_c = 0.4$ which from one side excludes small, partly numerically incorrect, values of λ and on the other side gives sufficiently large values of N_λ. Here we have $\nu = 0.49 \pm 02$ corresponding to the fractal dimension $d = 0.98 \pm 0.04$. Furthermore, for $0.4 \leq \lambda_c \leq 0.7$ we have a rather constant value $\nu \approx 0.5$ with $d_f \approx 1.0$. Of course, it would be interesting to extend this analysis to a larger size N of citation networks of various type and not only for Phys. Rev. We expect that the fractal Weyl law is a generic feature of citation networks.

Further studies of the citation network of Physical Review concern the properties of eigenvectors (different from the PageRank) associated to relatively large complex eigenvalues, the fractal Weyl law, the correlations between PageRank and CheiRank (see also subsection IV.C) and the notion of "ImpactRank" (Frahm *et al.*, 2014b). To define the ImpactRank one may ask the question how a paper influences or has been influenced by other papers. For this one considers an initial vector v_0, localized on a one node/paper. Then the modified Google matrix $\tilde{G} = \gamma\, G + (1-\gamma)\, v_0\, e^T$ (with a damping factor $\gamma \sim 0.5-0.9$) produces a "PageRank" v_f by the propagator $v_f = (1-\gamma)/(1-\gamma G)\, v_0$. In the vector v_f the leading nodes/papers have strongly influenced the initial paper represented in v_0. Doing the same for G^* one obtains a vector v_f^* where the leading papers have been influenced by the initial paper represented in v_0. This procedure has been applied to certain historically important papers (Frahm *et al.*, 2014b).

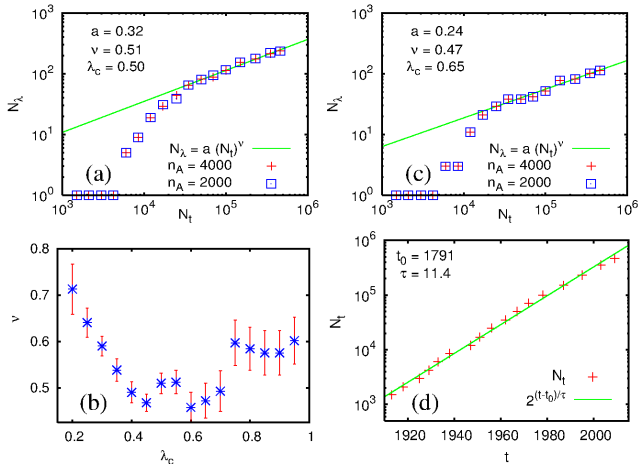

FIG. 46 (Color online) Data for the whole CNPR at different moments of time. Panel (a) (or (c)): shows the number N_λ of eigenvalues with $\lambda_c \leq \lambda \leq 1$ for $\lambda_c = 0.50$ (or $\lambda_c = 0.65$) versus the effective network size N_t where the nodes with publication times after a cut time t are removed from the network. The green/gray line shows the fractal Weyl law $N_\lambda = a\,(N_t)^\nu$ with parameters $a = 0.32 \pm 0.08$ ($a = 0.24 \pm 0.11$) and $\nu = 0.51 \pm 0.02$ ($b = 0.47 \pm 0.04$) obtained from a fit in the range $3 \times 10^4 \leq N_t < 5 \times 10^5$. The number N_λ includes both exactly determined invariant subspace eigenvalues and core space eigenvalues obtained from the Arnoldi method with double-precision (52 binary digits) for $n_A = 4000$ (red/gray crosses) and $n_A = 2000$ (blue/black squares). Panel (b): exponent b with error bars obtained from the fit $N_\lambda = a\,(N_t)^\nu$ in the range $3 \times 10^4 \leq N_t < 5 \times 10^5$ versus cut value λ_c. Panel (d): effective network size N_t versus cut time t (in years). The green/gray line shows the exponential fit $2^{(t-t_0)/\tau}$ with $t_0 = 1791 \pm 3$ and $\tau = 11.4 \pm 0.2$ representing the number of years after which the size of the network (number of papers published in all Physical Review journals) is effectively doubled. After (Frahm *et al.*, 2014b).

In summary, the results of this section show that the phenomenon of the Jordan error enhancement (12), induced by finite accuracy of computations with a finite number of digits, can be resolved by advanced numerical methods described above. Thus the accurate eigenvalues λ can be obtained even for the most difficult case of quasi-triangular matrices. We note that for other networks like WWW of UK universities, Wikipedia and Twitter the triangular structure of S is much less pronounced (see e.g. Fig. 1) that gives a reduction of Jordan blocks so that the Arnoldi method with double precision computes

accurate values of λ.

XIII. RANDOM MATRIX MODELS OF MARKOV CHAINS

A. Albert-Barabási model of directed networks

There are various preferential attachment models generating complex scale-free networks (see e.g. (Albert and Barabási , 2002; Dorogovtsev, 2010)). Such undirected networks are generated by the Albert-Barabási (AB) procedure (Albert and Barabási , 2000) which builds networks by an iterative process. Such a procedure has been generalized to generate directed networks in (Giraud *et al.*, 2009) with the aim to study properties of the Google matrix of such networks. The procedure is working as follows: starting from m nodes, at each step m links are added to the existing network with probability p, or m links are rewired with probability q, or a new node with m links is added with probability $1-p-q$. In each case the end node of new links is chosen with preferential attachment, i.e. with probability $(k_i+1)/\sum_j(k_j+1)$ where k_i is the total number of ingoing and outgoing links of node i. This mechanism generates directed networks having the small-world and scale-free properties, depending on the values of p and q. The results are averaged over N_r random realizations of the network to improve the statistics.

The studies (Giraud *et al.*, 2009) are done mainly for $m = 5$, $p = 0.2$ and two values of q corresponding to scale-free ($q = 0.1$) and exponential ($q = 0.7$) regimes of link distributions (see Fig. 1 in (Albert and Barabási , 2000) for undirected networks). For the generated directed networks at $q = 0.1$, one finds properties close to the behavior for the WWW with the cumulative distribution of ingoing links showing algebraic decay $P_c^{\rm in}(k) \sim 1/k$ and average connectivity $\langle k\rangle \approx 6.4$. For $q = 0.7$ one finds $P_c^{\rm in}(k) \sim \exp(-0.03k)$ and $\langle k\rangle \approx 15$. For outgoing links, the numerical data are compatible with an exponential decay in both cases with $P_c^{\rm out}(k) \sim \exp(-0.6k)$ for $q = 0.1$ and $P_c^{\rm out}(k) \sim \exp(-0.1k)$ for $q = 0.7$. It is found that small variations of parameters m, p, q near the chosen values do not qualitatively affect the properties of G matrix.

It is found that the eigenvalues of G for the AB model have one $\lambda = 1$ with all other $|\lambda_i| < 0.3$ at $\alpha = 0.85$ (see Fig. 1 in (Giraud *et al.*, 2009)). This distribution shows no significant modification with the growth of matrix size $2^{10} \leq N \leq 2^{14}$. However, the values of IPR ξ are growing with N for typical values $|\lambda| \sim 0.2$. This indicates a delocalization of corresponding eigenstates at large N. At the same time the PageRank probability is well described by the algebraic dependence $P \sim 1/K$ with ξ being practically independent of N.

These results for directed AB model network shows that it captures certain features of real directed networks, as e.g. a typical PageRank decay with the exponent $\beta \approx 1$. However, the spectrum of G in this model is characterized by a large gap between $\lambda = 1$ and other eigenvalues which have $\lambda \leq 0.35$ at $\alpha = 1$. This feature is in a drastic difference with spectra of such typical networks at WWW of universities, Wikipedia and Twitter (see Figs. 17,22,32). In fact the AB model has no subspaces and no isolated or weakly coupled communities. In this network all sites can be reached from a given site in a logarithmic number of steps that generates a large gap in the spectrum of Google matrix and a rapid relaxation to PageRank eigenstate. In real networks there are plenty of isolated or weakly coupled communities and the introduction of damping factor $\alpha < 1$ is necessary to have a single PageRank eigenvalue at $\lambda = 1$. Thus the results obtained in (Giraud *et al.*, 2009) show that the AB model is not able to capture the important spectral features of real networks.

Additional studies in (Giraud *et al.*, 2009) analyzed the model of a real WWW university network with rewiring procedure of links, which consists in randomizing the links of the network keeping fixed the number of links at any given node. Starting from a single network, this creates an ensemble of randomized networks of same size, where each node has the same number of ingoing and outgoing links as for the original network. The spectrum of such randomly rewired networks is also characterized by a large gap in the spectrum of G showing that rewiring destroys the communities existing in original networks. The spectrum and eigenstate properties are studied in the related work on various real networks of moderate size $N < 2 \times 10^4$ which have no spectral gap (Georgeot *et al.*, 2010).

B. Random matrix models of directed networks

Above we saw that the standard models of scale-free networks are not able to reproduce the typical properties of spectrum of Google matrices of real large scale networks. At the same time we believe that it is important to find realistic matrix models of WWW and other networks. Here we discuss certain results for certain random matrix models of G.

Analytical and numerical studies of random unistochastic or orthostochastic matrices of size $N = 3$ and 4 lead to triplet and cross structures in the complex eigenvalue spectra (Zyczkowski *et al.*, 2003) (see also Fig. 18). However, the size of such matrices is too small.

Here we consider other examples of random matrix models of Perron-Frobenius operators characterized by non-negative matrix elements and column sums normalized to unity. We call these models Random Perron-Frobenius Matrices (RPFM). A number of RPFM, with arbitrary size N, can be constructed by drawing N^2 independent matrix elements $0 \leq G_{ij} \leq 1$ from a given distribution $p(G_{ij})$ with finite variance $\sigma^2 = \langle G_{ij}^2\rangle - \langle G_{ij}\rangle^2$ and normalizing the column sums to unity (Frahm *et al.*, 2014b). The average matrix $\langle G_{ij}\rangle = 1/N$ is just a pro-

jector on the vector e (with unity entries on each node, see also Sec. XII.A) and has the two eigenvalues $\lambda_1 = 1$ (of multiplicity 1) and $\lambda_2 = 0$ (of multiplicity $N-1$). Using an argument of degenerate perturbation theory on $\delta G = G - \langle G \rangle$ and known results on the eigenvalue density of non-symmetric random matrices (Akemann *et al.*, 2011; Guhr *et al.*, 1998; Mehta, 2004) one finds that an arbitrary realization of G has the leading eigenvalue $\lambda_1 = 1$ and the other eigenvalues are uniformly distributed on the complex unit circle of radius $R = \sqrt{N}\sigma$ (see Fig. 47).

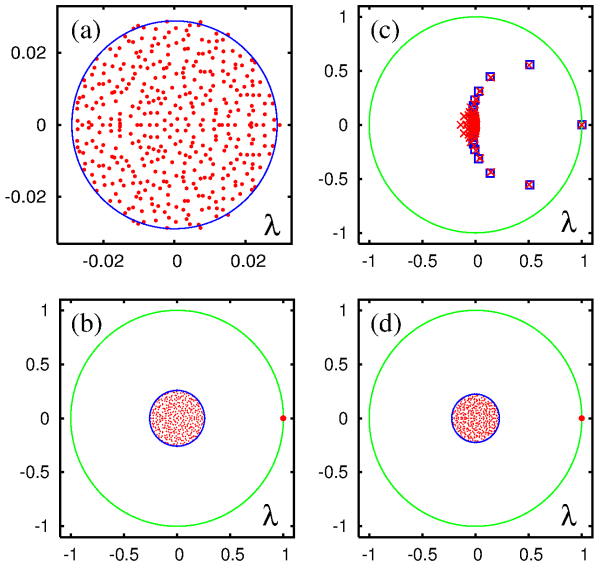

FIG. 47 (Color online) Panel (a) shows the spectrum (red/gray dots) of one realization of a full uniform RPFM with dimension $N = 400$ and matrix elements uniformly distributed in the interval $[0, 2/N]$; the blue/black circle represents the theoretical spectral border with radius $R = 1/\sqrt{3N} \approx 0.02887$. The unit eigenvalue $\lambda = 1$ is not shown due to the zoomed presentation range. Panel (c) shows the spectrum of one realization of triangular RPFM (red/gray crosses) with non-vanishing matrix elements uniformly distributed in the interval $[0, 2/(j-1)]$ and a triangular matrix with non-vanishing elements $1/(j-1)$ (blue/black squares); here $j = 2, 3, \ldots, N$ is the index-number of non-empty columns and the first column with $j = 1$ corresponds to a dangling node with elements $1/N$ for both triangular cases. Panels $(b), (d)$ show the complex eigenvalue spectrum (red/gray dots) of a sparse RPFM with dimension $N = 400$ and $Q = 20$ non-vanishing elements per column at random positions. Panel (b) (or (d)) corresponds to the case of uniformly distributed non-vanishing elements in the interval $[0, 2/Q]$ (constant non-vanishing elements being $1/Q$); the blue/black circle represents the theoretical spectral border with radius $R = 2/\sqrt{3Q} \approx 0.2582$ ($R = 1/\sqrt{Q} \approx 0.2236$). In panels $(b), (d)$ $\lambda = 1$ is shown by a larger red dot for better visibility. The unit circle is shown by green/gray curve (panels $(b), (c), (d)$). After (Frahm *et al.*, 2014b).

Choosing different distributions $p(G_{ij})$ one obtains different variants of the model (Frahm *et al.*, 2014b), for example $R = 1/\sqrt{3N}$ using a full matrix with uniform $G_{ij} \in [0, 2/N]$. Sparse models with $Q \ll N$ non-vanishing elements per column can be modeled by a distribution where the probability of $G_{ij} = 0$ is $1 - Q/N$ and for non-zero G_{ij} (either uniform in $[0, 2/Q]$ or constant $1/Q$) is Q/N leading to $R = 2/\sqrt{3Q}$ (for uniform non-zero elements) or $R = 1/\sqrt{Q}$ (for constant non-zero elements). The circular eigenvalue density with these values of R is also very well confirmed by numerical simulations in Fig. 47. Another case is a power law $p(G) = D/(1 + aG)^{-b}$ (for $0 \leq G \leq 1$) with D and a to be determined by normalization and the average $\langle G_{ij} \rangle = 1/N$. For $b > 3$ this case is similar to a full matrix with $R \sim 1/\sqrt{N}$. However for $2 < b < 3$ one finds that $R \sim N^{1-b/2}$.

The situation changes when one imposes a triangular structure on G in which case the complex spectrum of $\langle G \rangle$ is already quite complicated and, due to non-degenerate perturbation theory, close to the spectrum of G with modest fluctuations, mostly for the smallest eigenvalues (Frahm *et al.*, 2014b). Following the above discussion about triangular networks (with $G_{ij} = 0$ for $i \geq j$) we also study numerically a triangular RPFM where for $j \geq 2$ and $i < j$ the matrix elements G_{ij} are uniformly distributed in the interval $[0, 2/(j-1)]$ and for $i \geq j$ we have $G_{ij} = 0$. Then the first column is empty, that means it corresponds to a dangling node and it needs to be replaced by $1/N$ entries. For the triangular RPFM the situation changes completely since here the average matrix $\langle G_{ij} \rangle = 1/(j-1)$ (for $i < j$ and $j \geq 2$) has already a nontrivial structure and eigenvalue spectrum. Therefore the argument of degenerate perturbation theory which allowed to apply the results of standard full non-symmetric random matrices does not apply here. In Fig. 47 one clearly sees that for $N = 400$ the spectra for one realization of a triangular RPFM and its average are very similar for the eigenvalues with large modulus but both do not have at all a uniform circular density in contrast to the RPRM models without the triangular constraint discussed above. For the triangular RPFM the PageRank behaves as $P(K) \sim 1/K$ with the ranking index K being close to the natural order of nodes $\{1, 2, 3, \ldots\}$ that reflects the fact that the node 1 has the maximum of $N-1$ incoming links etc.

The above results show that it is not so simple to propose a good random matrix model which captures the generic spectral features of real directed networks. We think that investigations in this direction should be continued.

C. Anderson delocalization of PageRank?

The phenomenon of Anderson localization of electron transport in disordered materials (Anderson , 1958) is now a well-known effect studied in detail in physics (see e.g. (Evers and Mirlin , 2008)). In one and two dimensions even a small disorder leads to an exponential localization of electron diffusion that corresponds to an insu-

lating phase. Thus, even if a classical electron dynamics is diffusive and delocalized over the whole space, the effects of quantum interference generates a localization of all eigenstates of the Schödinger equation. In higher dimensions a localization is preserved at a sufficiently strong disorder, while a delocalized metallic phase appears for a disorder strength being smaller a certain critical value dependent on the Fermi energy of electrons. This phenomenon is rather generic and we can expect that a somewhat similar delocalization transition can appear in the small-world networks.

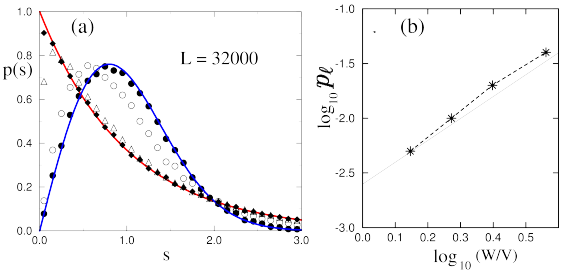

FIG. 48 (Color online) (a) The red/gray and blue/black curves represent the Poisson and Wigner surmise distributions. Diamonds, triangles, circles and black disks represent respectively the level spacing statistics $p(s)$ at $W/V = 4, 3, 2, 1$; $p_\ell = 0.02$, $L = 32000$; averaging is done over 60 network realizations. (b) Stars give dependence of p_ℓ on a disorder strength W/V at the critical point when $\eta_\ell(W, p_\ell) = 0.8$, and $p_\ell = 0.005, 0.01, 0.02, 0.04$ at fixed $L = 8000$; the straight line corresponds to $p_\ell = p_c = 1/4\ell_1 \approx (W/V)^2/400$; the dashed curve is drown to adapt an eye. After (Chepelianskii and Shepelyansky , 2001).

Indeed, it is useful to consider the $1D$ Anderson model on a ring with a certain number of shortcut links, described by the Schödinger equation

$$\epsilon_n\psi_n + V(\psi_{n+1} + \psi_{n-1}) + V\sum_S(\psi_{n+S} + \psi_{n-S}) = E\psi_n \ , \tag{15}$$

where ϵ_n are random on site energies homogeneously distributed within the interval $-W/2 \le \epsilon_n \le W/2$, and V is the hopping matrix element. The sum over S is taken over randomly established shortcuts from a site n to any other random site of the network. The number of such shortcuts is $S_{\rm tot} = p_\ell L$, where L is the total number of sites on a ring and p_ℓ is the density of shortcut links. This model had been introduced in (Chepelianskii and Shepelyansky , 2001). The numerical study, reported there, showed that the level-spacing statistics $p(s)$ for this model has a transition from the Poisson distribution $p_{\rm Pois}(s) = \exp(-s)$, typical for the Anderson localization phase, to the Wigner surmise distribution $p_{\rm Wig}(s) = \pi s/2 \exp(-\pi s^2/4)$, typical for the Anderson metallic phase (Evers and Mirlin , 2008; Guhr *et al.*, 1998). The numerical diagonalization was done via the Lanczos algorithm for the sizes up to $L = 32000$ and the typical parameter range $0.005 \le p_\ell < 0.1$ and $1 \le W/V \le 4$. An example, of the variation of $p_\ell(s)$ with a decrease of W/V is shown in Fig. 48(a). We see that the Wigner surmise provides a good description of the numerical data at $W/V = 1$, when the maximal localization length $\ell_1 \approx 96(V/W)^2 \approx 96$ in the 1D Anderson model (see e.g. (Evers and Mirlin , 2008)) is much smaller than the system size L.

To identify a transition from one limiting case $p_{\rm Pois}(s)$ to another $p_{\rm Wig}(s)$ it is convenient to introduce the parameter $\eta_s = \int_0^{s_0}(p(s) - p_{\rm Wig}(s))ds / \int_0^{s_0}(p_{\rm Pois}(s) - p_{\rm Wig}(s))ds$, where $s_0 = 0.4729...$ is the intersection point of $p_{\rm Pois}(s)$ and $p_{\rm Wig}(s)$. In this way η_s varies from 1 (for $p(s) = p_{\rm Pois}(s)$) to 0 (for $p(s) = p_{\rm Wig}(s)$) (see e.g. (Shepelyansky , 2001)). From the variation of η_s with system parameters and size L, the critical density $p_\ell = p_c$ can be determined by the condition $\eta_s(p_c, W/V) = \eta_c = 0.8 = const.$ being independent of L. The obtained dependence of p_c on W/V obtained at a fixed critical point $\eta_c = 0.8$ is shown in Fig. 48(b). The Anderson delocalization transition takes place when the density of shortcuts becomes larger than a critical density $p_\ell > p_c \approx 1/(4\ell_1)$ where $\ell_1 \approx 96(V/W)^2$ is the length of Anderson localization in $1D$. A simple physical interpretation of this result is that the delocalization takes place when the localization length ℓ_1 becomes larger than a typical distance $1/(4p_\ell)$ between shortcuts. The further studies of time evolution of wave function $\psi_n(t)$ and IPR ξ variation also confirmed the existence of quantum delocalization transition on this quantum small-world network (Giraud *et al.*, 2005).

Thus the results obtained for the quantum small-world networks (Chepelianskii and Shepelyansky , 2001; Giraud *et al.*, 2005) show that the Anderson transition can take place in such systems. However, the above model represents an undirected network corresponding to a symmetric matrix with a real spectrum while the typical directed networks are characterized by asymmetric matrix G and complex spectrum. The possibility of existence of localized states of G for WWW networks was also discussed by (Perra *et al.*, 2009) but the fact that in a typical case the spectrum of G is complex has not been analyzed in detail.

Above we saw certain indications on a possibility of Anderson type delocalization transition for eigenstates of the G matrix. Our results clearly show that certain eigenstates in the core space are exponentially localized (see e.g. Fig 19(b)). Such states are localized only on a few nodes touching other nodes of network only by an exponentially small tail. A similar situation would appear in the 1D Anderson model if an absorption would be introduced on one end of the chain. Then the eigenstates located far away from this place would feel this absorption only by exponentially small tails so that the imaginary part of the eigenenergy would have for such far away states only an exponentially small imaginary part. It is natural to expect that such localization can be destroyed by some parameter variation. Indeed, certain eigenstates with $|\lambda| < 1$ for the directed network of the AB model have IPR ξ growing with the matrix

size N (see Sec. XIII.A and (Giraud *et al.*, 2009)) even if for the PageRank the values of ξ remain independent of N. The results for the Ulam network from Figs. 13, 14 provide an example of directed network where the PageRank vector becomes delocalized when the damping factor is decreased from $\alpha = 0.95$ to 0.85 (Zhirov *et al.*, 2010). This example demonstrates a possibility of PageRank delocalization but a deeper understanding of the conditions required for such a phenomenon to occur are still lacking. The main difficulty is an absence of well established random matrix models which have properties similar to the available examples of real networks.

Indeed, for Hermitian and unitary matrices the theories of random matrices, mesoscopic systems and quantum chaos allow to capture main universal properties of spectra and eigenstates (Akemann *et al.*, 2011; Evers and Mirlin , 2008; Guhr *et al.*, 1998; Haake, 2010; Mehta, 2004). For asymmetric Google matrices the spectrum is complex and at the moment there are no good random matrix models which would allow to perform analytical analysis of various parameter dependencies. It is possible that non-Hermitian Anderson models in $1D$, which naturally generates a complex spectrum and may have delocalized eigenstates, will provide new insights in this direction (Goldsheid and Khoruzhenko , 1998).

XIV. OTHER EXAMPLES OF DIRECTED NETWORKS

In this section we discuss additional examples of real directed networks.

A. Brain neural networks

In 1958 John von Neumann traced first parallels between architecture of the computer and the brain (von Neumann , 1958). Since that time computers became an unavoidable element of the modern society forming a computer network connected by the WWW with about 4×10^9 indexed web pages spread all over the world (see e.g. http://www.worldwidewebsize.com/). This number starts to become comparable with 10^{10} neurons in a human brain where each neuron can be viewed as an independent processing unit connected with about 10^4 other neurons by synaptic links (see e.g. (Sporns , 2007)). About 20% of these links are unidirectional (Felleman and van Essen , 1991) and hence the brain can be viewed as a directed network of neuron links. At present, more and more experimental information about neurons and their links becomes available and the investigations of properties of neuronal networks attract an active interest (see e.g. (Bullmore and Sporns , 2009; Zuo *et al.*, 2012)). The fact that enormous sizes of WWW and brain networks are comparable gives an idea that the Google matrix analysis should find useful application in brain science as it is the case of WWW.

First applications of methods of Google matrix methods to brain neural networks was done in (Shepelyansky and Zhirov , 2010b) for a large-scale thalamocortical model (Izhikevich and Edelman , 2008) based on experimental measures in several mammalian species. The model spans three anatomic scales. (i) It is based on global (white-matter) thalamocortical anatomy obtained by means of diffusion tensor imaging of a human brain. (ii) It includes multiple thalamic nuclei and six-layered cortical microcircuitry based on in vitro labeling and three-dimensional reconstruction of single neurons of cat visual cortex. (iii) It has 22 basic types of neurons with appropriate laminar distribution of their branching dendritic trees. According to (Izhikevich and Edelman , 2008) the model exhibits behavioral regimes of normal brain activity that were not explicitly built-in but emerged spontaneously as the result of interactions among anatomical and dynamic processes.

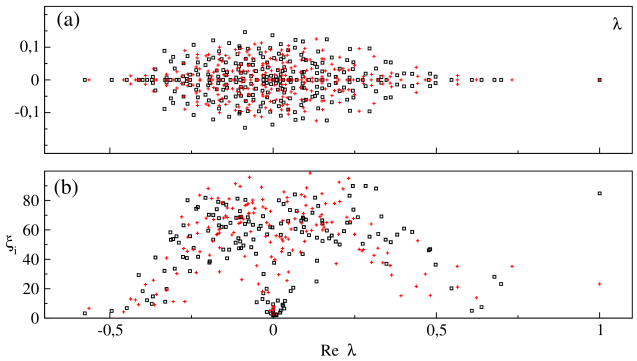

FIG. 49 (Color online) (a) Spectrum of eigenvalues λ for the Google matrices G and G^* at $\alpha = 0.85$ for the neural network of *C.elegans* (black and red/gray symbols). (b) Values of IPR ξ_i of eigenvectors ψ_i are shown as a function of corresponding $Re\lambda$ (same colors). After (Kandiah and Shepelyansky , 2014a).

The model studied in (Shepelyansky and Zhirov , 2010b) contains $N = 10^4$ neuron with $N_\ell = 1960108$. The obtained results show that PageRank and CheiRank vectors have rather large ξ being comparable with the whole network size at $\alpha = 0.85$. The corresponding probabilities have very flat dependence on their indexes showing that they are close to a delocalized regime. We attribute these features to a rather large number of links per node $\zeta \approx 196$ being even larger than for the Twitter network. At the same time the PageRank-CheiRank correlator is rather small $\kappa = -0.065$. Thus this network is structured in such a way that functions related to order signals (outgoing links of CheiRank) and signals bringing orders (ingoing links of PageRank) are well separated and independent of each other as it is the case for the Linux Kernel software architecture. The spectrum of G has a gapless structure showing that long living excitations can exist in this neuronal network.

Of course, model systems of neural networks can provide a number of interesting insights but it is much more

important to study examples of real neural networks. In (Kandiah and Shepelyansky , 2014a) such an analysis is performed for the neural network of *C.elegans* (worm). The full connectivity of this directed network is known and well documented at WormAtlas (Altun *et al.*, 2012). The number of linked neurons (nodes) is $N = 279$ with the number of synaptic connections and gap junctions (links) between them being $N_\ell = 2990$.

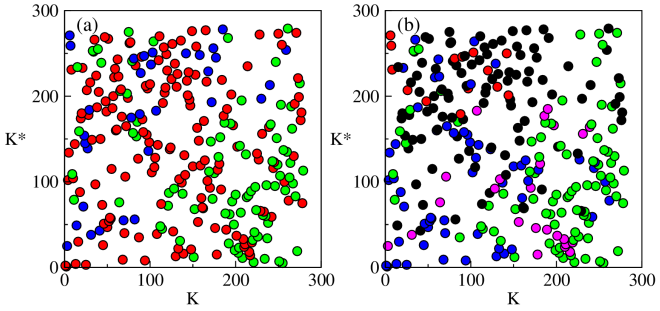

FIG. 50 (Color online) PageRank - CheiRank plane (K, K^*) showing distribution of neurons according to their ranking. (a): soma region coloration - head (red/gray), middle (green/light gray), tail (blue/dark gray). (b): neuron type coloration - sensory (red/gray), motor (green/light gray), interneuron (blue/dark gray), polymodal (purple/light-dark gray) and unknown (black). The classifications and colors are given according to WormAtlas (Altun *et al.*, 2012). After (Kandiah and Shepelyansky , 2014a).

The Google matrix G of *C.elegans* is constructed using the connectivity matrix elements $S_{ij} = S_{\rm syn,ij} + S_{\rm gap,ij}$, where $S_{\rm syn}$ is an asymmetric matrix of synaptic links whose elements are 1 if neuron j connects to neuron i through a chemical synaptic connection and 0 otherwise. The matrix part $S_{\rm gap}$ is a symmetric matrix describing gap junctions between pairs of cells, $S_{\rm gap,ij} = S_{\rm gap,ji} = 1$ if neurons i and j are connected through a gap junction and 0 otherwise. Then the matrices G and G^* are constructed following the standard rule (1) at $\alpha = 0.85$. The connectivity properties of this network are similar to those of WWW of Cambridge and Oxford with approximately the same number of links per node.

The spectra of G and G^* are shown in Fig. 49 with corresponding IPR values of eigenstates. The imaginary part of λ is relatively small $|{\rm Im}(\lambda)| < 0.2$ due to a large fraction of symmetric links. The second by modulus eigenvalues are $\lambda_2 = 0.8214$ for G and $\lambda_2 = 0.8608$ for G^*. Thus the network relaxation time $\tau = 1/|\ln \lambda_2|$ is approximately $5, 6.7$ iterations of G, G^*. Certain IPR values ξ_i of eigenstates of G, G^* have rather large $\xi \approx N/3$ while others have ξ located only on about ten nodes.

We have a large value $\xi \approx 85$ for PageRank and a more moderate value $\xi \approx 23$ for CheiRank vectors. Here we have the algebraic decay exponents being $\beta \approx 0.33$ for $P(K)$ and $\beta \approx 0.50$ for $P^*(K^*)$. Of course, the network size is not large and these values are only approximate. However, they indicate an interchange between PageRank and CheiRank showing importance of outgoing links. It is possible that such an inversion is related to a significant importance of outgoing links in neural systems: in a sense such links transfer orders, while ingoing links bring instructions to a given neuron from other neurons. The correlator $\kappa = 0.125$ is small and thus, the network structure allows to perform a control of information flow in a more efficient way without interference of errors between orders and executions. We saw already in Sec. VII.A that such a separation of concerns emerges in software architecture. It seems that the neural networks also adopt such a structure.

We note that a somewhat similar situation appears for networks of Business Process Management where *Principals* of a company are located at the top CheiRank position while the top PageRank positions belong to company *Contacts* (Abel and Shepelyansky , 2011). Indeed, a case study of a real company structure analyzed in (Abel and Shepelyansky , 2011) also stress the importance of company managers who transfer orders to other structural units. For this network the correlator is also small being $\kappa = 0.164$. We expect that brain neural networks may have certain similarities with company organization.

Each neuron i belongs to two ranks K_i and K_i^* and it is convenient to represent the distribution of neurons on PageRank-CheiRank plane (K, K^*) shown in Fig. 50. The plot confirms that there are little correlations between both ranks since the points are scattered over the whole plane. Neurons ranked at top K positions of PageRank have their soma located mainly in both extremities of the worm (head and tail) showing that neurons in those regions have important connections coming from many other neurons which control head and tail movements. This tendency is even more visible for neurons at top K^* positions of CheiRank but with a preference for head and middle regions. In general, neurons, that have their soma in the middle region of the worm, are quite highly ranked in CheiRank but not in PageRank. The neurons located at the head region have top positions in CheiRank and also PageRank, while the middle region has some top CheiRank indexes but rather large indexes of PageRank (Fig. 50 (a)). The neuron type coloration (Fig. 50 (b)) also reveals that sensory neurons are at top PageRank positions but at rather large CheiRank indexes, whereas in general motor neurons are in the opposite situation.

Top nodes of PageRank and CheiRank favor important signal relaying neurons such as AVA and AVB that integrate signals from crucial nodes and in turn pilot other crucial nodes. Neurons $AVAL, AVAR$, $AVBL, AVBR$ and $AVEL, AVER$ are considered to belong to the rich club analyzed in (Towlson *et al.*, 2013). The top neurons in 2DRank are AVAL, AVAR, AVBL, AVBR, PVCR that corresponds to a dominance of interneurons. More details can be found in (Kandiah and Shepelyansky , 2014a).

The technological progress allows to obtain now more and more detailed information about neural networks (see e.g. (Bullmore and Sporns , 2009; Towlson *et al.*, 2013; Zuo *et al.*, 2012)) even if it is not easy to get infor-

mation about link directions. In view of that we expect that the methods of directed network analysis described here will find useful future applications for brain neural networks.

B. Google matrix of DNA sequences

The approaches of Markov chains and Google matrix can be also efficiently used for analysis of statistical properties of DNA sequences. The data sets are publicly available at (Ensemble Genome database, 2011). The analysis of Poincaré recurrences in these DNA sequences (Frahm and Shepelyansky , 2012c) shows their similarities with the statistical properties of recurrences for dynamical trajectories in the Chirikov standard map and other symplectic maps (Frahm and Shepelyansky , 2010). Indeed, a DNA sequence can be viewed as a long symbolic trajectory and hence, the Google matrix, constructed from it, highlights the statistical features of DNA from a new viewpoint.

An important step in the statistical analysis of DNA sequences was done in (Mantegna *et al.*, 1995) applying methods of statistical linguistics and determining the frequency of various words composed of up to 7 letters. A first order Markovian models have been also proposed and briefly discussed in this work. The Google matrix analysis provides a natural extension of this approach. Thus the PageRank eigenvector gives most frequent words of given length. The spectrum and eigenstates of G characterize the relaxation processes of different modes in the Markov process generated by a symbolic DNA sequence. Thus the comparison of word ranks of different species allows to identify their proximity.

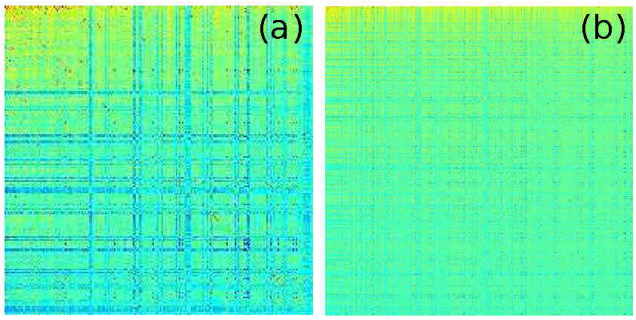

FIG. 51 (Color online) DNA Google matrix of Homo sapiens (HS) constructed for words of 6-letters length. Matrix elements $G_{KK'}$ are shown in the basis of PageRank index K (and K'). Here, x and y axes show K and K' within the range $1 \le K, K' \le 200$ (a) and $1 \le K, K' \le 1000$ (b). The element G_{11} at $K = K' = 1$ is placed at top left corner. Color marks the amplitude of matrix elements changing from blue/black for minimum zero value to red/gray at maximum value. After (Kandiah and Shepelyansky , 2013).

The statistical analysis is done for DNA sequences of the species: Homo sapiens (HS, human), Canis familiaris (CF, dog), Loxodonta africana (LA, elephant), Bos Taurus (bull, BT), Danio rerio (DR, zebrafish) (Kandiah and Shepelyansky , 2013). For HS DNA sequences are represented as a single string of length $L \approx 1.5 \cdot 10^{10}$ base pairs (bp) corresponding to 5 individuals. Similar data are obtained for BT ($2.9 \cdot 10^9$ bp), CF ($2.5 \cdot 10^9$ bp), LA ($3.1 \cdot 10^9$ bp), DR ($1.4 \cdot 10^9$ bp). All strings are composed of 4 letters A, G, G, T and undetermined letter N_l. The strings can be found from (Kandiah and Shepelyansky , 2013).

For a given sequence we fix the words W_k of m letters length corresponding to the number of states $N = 4^m$. We consider that there is a transition from a state j to state i inside this basis N when we move along the string from left to right going from a word W_k to a next word W_{k+1}. This transition adds one unit in the transition matrix element $T_{ij} \to T_{ij} + 1$. The words with letter N_l are omitted, the transitions are counted only between nearby words not separated by words with N_l. There are approximately $N_t \approx L/m$ such transitions for the whole length L since the fraction of undetermined letters N_l is small. Thus we have $N_t = \sum_{i,j=1}^{N} T_{ij}$. The Markov matrix of transitions S_{ij} is obtained by normalizing matrix elements in such a way that their sum in each column is equal to unity: $S_{ij} = T_{ij} / \sum_i T_{ij}$. If there are columns with all zero elements (dangling nodes) then zeros of such columns are replaced by $1/N$. Then the Google matrix G is constructed from S by the standard rule (1). It is found that the spectrum of G has a significant gap and a variation of α in a range $(0.5, 1)$ does not affect significantly the PageRank probability. Thus all DNA results are shown at $\alpha = 1$.

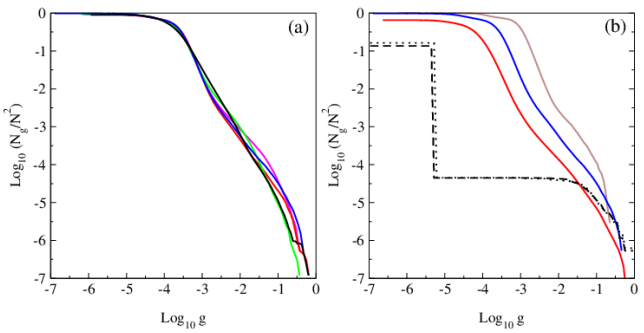

FIG. 52 (Color online) Integrated fraction N_g/N^2 of Google matrix elements with $G_{ij} > g$ as a function of g. (a) Various species with 6-letters word length: elephant LA (green), zebrafish DR(black), dog CF (red), bull BT (magenta), and Homo sapiens HS (blue) (from left to right at $y = -5.5$). (b) Data for HS sequence with words of length $m = 5$ (brown), 6 (blue), 7 (red) (from right to left at $y = -2$); for comparison black dashed and dotted curves show the same distribution for the WWW networks of Universities of Cambridge and Oxford in 2006 respectively. After (Kandiah and Shepelyansky , 2013).

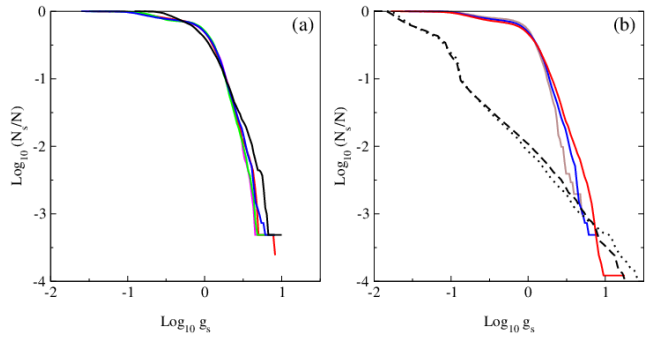

FIG. 53 (Color online) Integrated fraction N_s/N of sum of ingoing matrix elements with $\sum_{j=1}^{N} G_{i,j} \geq g_s$. Panels (a) and (b) show the same cases as in Fig. 52 in same colors. The dashed and dotted curves are shifted in x-axis by one unit left to fit the figure scale. After (Kandiah and Shepelyansky , 2013).

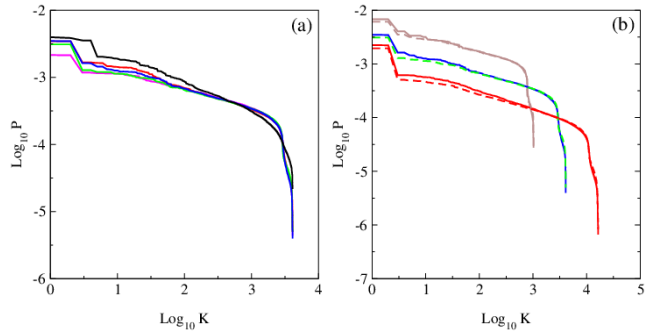

FIG. 54 (Color online) Dependence of PageRank probability $P(K)$ on PageRank index K. (a) Data for different species for word length of 6-letters: zebrafish DR (black), dog CF (red), Homo sapiens HS (blue), elephant LA (green) and bull BT (magenta) (from top to bottom at $x = 1$). (b) Data for HS (full curve) and LA (dashed curve) for word length $m = 5$ (brown), 6 (blue/green), 7 (red) (from top to bottom at $x = 1$). After (Kandiah and Shepelyansky , 2013).

The image of matrix elements $G_{KK'}$ is shown in Fig. 51 for HS with $m = 6$. We see that almost all matrix is full that is drastically different from the WWW and other networks considered above. The analysis of statistical properties of matrix elements G_{ij} shows that their integrated distribution follows a power law as it is seen in Fig. 52. Here N_g is the number of matrix elements of the matrix G with values $G_{ij} > g$. The data show that the number of nonzero matrix elements G_{ij} is very close to N^2. The main fraction of elements has values $G_{ij} \leq 1/N$ (some elements $G_{ij} < 1/N$ since for certain j there are many transitions to some node i' with $T_{i'j} \gg N$ and e.g. only one transition to other i'' with $T_{i''j} = 1$). At the same time there are also transition elements G_{ij} with large values whose fraction decays in an algebraic law $N_g \approx AN/g^{\nu-1}$ with some constant A and an exponent ν. The fit of numerical data in the range $-5.5 < \log_{10} g < -0.5$ of algebraic decay gives for $m = 6$: $\nu = 2.46 \pm 0.025$ (BT), 2.57 ± 0.025 (CF), 2.67 ± 0.022 (LA), 2.48 ± 0.024 (HS), 2.22 ± 0.04 (DR). For HS case we find $\nu = 2.68 \pm 0.038$ at $m = 5$ and $\nu = 2.43 \pm 0.02$ at $m = 7$ with the average $A \approx 0.003$ for $m = 5, 6, 7$. There are visible oscillations in the algebraic decay of N_g with g but in global we see that on average all species are well described by a universal decay law with the exponent $\nu \approx 2.5$. For comparison we also show the distribution N_g for the WWW networks of University of Cambridge and Oxford in year 2006. We see that in these cases the distribution N_g has a very short range in which the decay is at least approximately algebraic ($-5.5 < \log_{10}(N_g/N^2) < -6$). In contrast to that for the DNA sequences we have a large range of algebraic decay.

Since in each column we have the sum of all elements equal to unity we can say that the differential fraction $dN_g/dg \propto 1/g^{\nu}$ gives the distribution of outgoing matrix elements which is similar to the distribution of outgoing links extensively studied for the WWW networks. Indeed, for the WWW networks all links in a column are considered to have the same weight so that these matrix elements are given by an inverse number of outgoing links with the decay exponent $\nu \approx 2.7$. Thus, the obtained data show that the distribution of DNA matrix elements is similar to the distribution of outgoing links in the WWW networks. Indeed, for outgoing links of Cambridge and Oxford networks the fit of numerical data gives the exponents $\nu = 2.80 \pm 0.06$ (Cambridge) and 2.51 ± 0.04 (Oxford).

As discussed above, on average the probability of PageRank vector is proportional to the number of ingoing links that works satisfactory for sparse G matrices. For DNA we have a situation where the Google matrix is almost full and zero matrix elements are practically absent. In such a case an analogue of number of ingoing links is the sum of ingoing matrix elements $g_s = \sum_{j=1}^{N} G_{ij}$. The integrated distribution of ingoing matrix elements with the dependence of N_s on g_s is shown in Fig. 53. Here N_s is defined as the number of nodes with the sum of ingoing matrix elements being larger than g_s. A significant part of this dependence, corresponding to large values of g_s and determining the PageRank probability decay, is well described by a power law $N_s \approx BN/g_s^{\mu-1}$. The fit of data at $m = 6$ gives $\mu = 5.59 \pm 0.15$ (BT), 4.90 ± 0.08 (CF), 5.37 ± 0.07 (LA), 5.11 ± 0.12 (HS), 4.04 ± 0.06 (DR). For HS case at $m = 5, 7$ we find respectively $\mu = 5.86 \pm 0.14$ and 4.48 ± 0.08. For HS and other species we have an average $B \approx 1$.

For WWW one usually have $\mu \approx 2.1$. Indeed, for the ingoing matrix elements of Cambridge and Oxford networks we find respectively the exponents $\mu = 2.12 \pm 0.03$ and 2.06 ± 0.02 (see curves in Fig. 53). For ingoing links distribution of Cambridge and Oxford networks we obtain respectively $\mu = 2.29 \pm 0.02$ and $\mu = 2.27 \pm 0.02$ which are close to the usual WWW value $\mu \approx 2.1$. In contrast the exponent μ for DNA Google matrix elements

gets significantly larger value $\mu \approx 5$. This feature marks a significant difference between DNA and WWW networks.

The PageRank vector can be obtained by a direct diagonalization. The dependence of probability P on index K is shown in Fig. 54 for various species and different word length m. The probability $P(K)$ describes the steady state of random walks on the Markov chain and thus it gives the frequency of appearance of various words of length m in the whole sequence L. The frequencies or probabilities of words appearance in the sequences have been obtained in (Mantegna *et al.*, 1995) by a direct counting of words along the sequence (the available sequences L were shorted at that times). Both methods are mathematically equivalent and indeed our distributions $P(K)$ are in good agreement with those found in (Mantegna *et al.*, 1995) even if now we have a significantly better statistics.

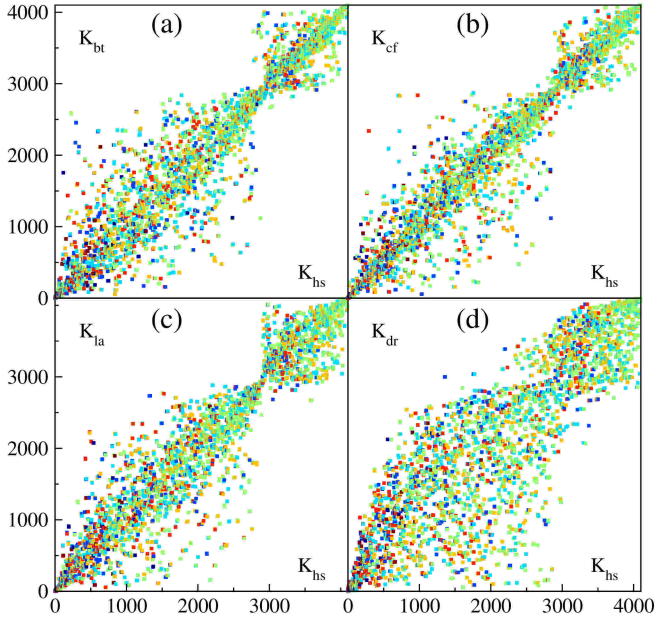

FIG. 55 (Color online) PageRank proximity $K - K$ plane diagrams for different species in comparison with Homo sapiens: (a) x-axis shows PageRank index $K_{hs}(i)$ of a word i and y-axis shows PageRank index of the same word i with $K_{bt}(i)$ of bull, (b) $K_{cf}(i)$ of dog, (c) $K_{la}(i)$ of elephant and (d) $K_{dr}(i)$ of zebrafish; here the word length is $m = 6$. The colors of symbols marks the purine content in a word i (fractions of letters A or G in any order); the color varies from red/gray at maximal content, via brown, yellow, green, light blue, to blue/black at minimal zero content. After (Kandiah and Shepelyansky , 2013).

The decay of P with K can be approximately described by a power law $P \sim 1/K^{\beta}$. Thus for example for HS sequence at $m = 7$ we find $\beta = 0.357 \pm 0.003$ for the fit range $1.5 \leq \log_{10} K \leq 3.7$ that is rather close to the exponent found in (Mantegna *et al.*, 1995). Since on average the PageRank probability is proportional to the number of ingoing links, or the sum of ingoing matrix elements of G, one has the relation between the exponent of PageRank β and exponent of ingoing links (or matrix elements): $\beta = 1/(\mu - 1)$. Indeed, for the HS DNA case at $m = 7$ we have $\mu = 4.48$ that gives $\beta = 0.29$ being close to the above value of $\beta = 0.357$ obtained from the direct fit of $P(K)$ dependence. The agreement is not so perfect since there is a visible curvature in the log-log plot of N_s vs g_s and also since a small value of β gives a moderate variation of P that produces a reduction of accuracy of numerical fit procedure. In spite of this only approximate agreement we conclude that in global the relation between β and μ works correctly.

It is interesting to plot a PageRank index $K_s(i)$ of a given species s versus the index $K_{hs}(i)$ of HS for the same word i. For identical sequences one should have all points on diagonal, while the deviations from diagonal characterize the differences between species. The examples of such PageRank proximity $K - K$ diagrams are shown in Fig. 55 for words at $m = 6$. A visual impression is that CF case has less deviations from HS rank compared to BT and LA. The non-mammalian DR case has most strong deviations from HS rank.

The fraction of purine letters A or G in a word of $m = 6$ letters is shown by color in Fig. 55 for all words ranked by PageRank index K. We see that these letters are approximately homogeneously distributed over the whole range of K values. To determine the proximity between different species or different HS individuals we compute the average dispersion

$$\sigma(s_1, s_2) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\Big(K_{s_1}(i) - K_{s_2}(i)\Big)^2} \qquad (16)$$

between two species (individuals) s_1 and s_2. Comparing the words with length $m = 5, 6, 7$ we find that the scaling $\sigma \propto N$ works with a good accuracy (about 10% when N is increased by a factor 16). To represent the result in a form independent of m we compare the values of σ with the corresponding random model value σ_{rnd}. This value is computed assuming a random distribution of N points in a square $N \times N$ when only one point appears in each column and each line (e.g. at $m = 6$ we have $\sigma_{rnd} \approx 1673$ and $\sigma_{rnd} \propto N$). The dimensionless dispersion is then given by $\zeta(s_1, s_2) = \sigma(s_1, s_2)/\sigma_{rnd}$. From the ranking of different species we obtain the following values at $m = 6$: $\zeta(CF, BT) = 0.308$; $\zeta(LA, BT) = 0.324$, $\zeta(LA, CF) = 0.303$; $\zeta(HS, BT) = 0.246$, $\zeta(HS, CF) = 0.206$, $\zeta(HS, LA) = 0.238$; $\zeta(DR, BT) = 0.425$, $\zeta(DR, CF) = 0.414$, $\zeta(DR, LA) = 0.422$, $\zeta(DR, HS) = 0.375$ (other m have similar values). According to this statistical analysis of PageRank proximity between species we find that ζ value is minimal between CF and HS showing that these are two most similar species among those considered here. The comparison of two HS individuals gives the value $\zeta(HS1, HS2) = 0.031$ being significantly smaller then the proximity correlator between different species (Kandiah and Shepelyansky , 2012).

The spectrum of G is analyzed in detail in (Kandiah

and Shepelyansky , 2012). It is shown that it has a relatively large gap due to which there is a relatively rapid relaxation of probability of a random surfer to the PageRank values.

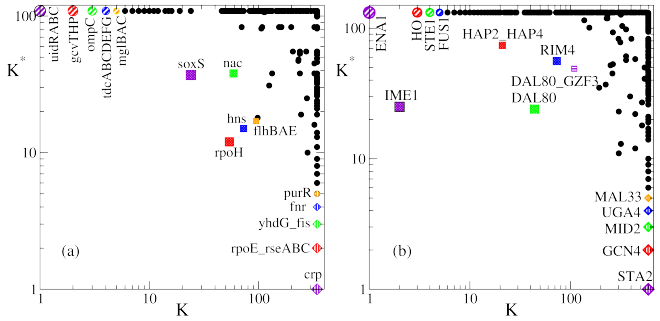

FIG. 56 (Color online) Distribution of nodes in the PageRank-CheiRank plane (K, K^*) for Escherichia Coli v1.1 (a), and Yeast (b) gene transcription networks on (network data are taken from (Milo *et al.*, 2002; Shen-Orr *et al.*, 2002) and (Alon, 2014)). The nodes with five top probability values of PageRank, CheiRank and 2DRank are labeled by their corresponding operon (node) names; they correspond to 5 lowest values of indexes K, K_2, K^*. After (Ermann *et al.*, 2012a).

C. Gene regulation networks

At present the analysis of gene transcription regulation networks and recovery of their control biological functions becomes an active research field of bioinformatics (see e.g. (Milo *et al.*, 2002)). Here, following (Ermann *et al.*, 2012a), we provide two simple examples of 2DRanking analysis for gene transcriptional regulation networks of Escherichia Coli ($N = 423$, $N_\ell = 519$ (Shen-Orr *et al.*, 2002)) and Yeast ($N = 690$, $N_\ell = 1079$ (Milo *et al.*, 2002)). In the construction of G matrix the outgoing links to all nodes in each column are taken with the same weight, $\alpha = 0.85$.

The distribution of nodes in PageRank-CheiRank plane is shown in Fig. 56. The top 5 nodes, with their operon names, are given there for indexes of PageRank K, CheiRank K^* and 2DRank K_2. This ranking selects operons with most high functionality in communication (K^*), popularity (K) and those that combines these both features (K_2). For these networks the correlator κ is close to zero ($\kappa = -0.0645$ for Escherichia Coli and $\kappa = -0.0497$ for Yeast, see Fig. 6)) that indicates the statistical independence between outgoing and ingoing links being quite similarly to the case of the PCN for the Linux Kernel. This may indicate that a slightly negative correlator κ is a generic property for the data flow network of control and regulation systems. A similar situation appears for networks of business process management and brain neural networks. Thus it is possible that the networks performing control functions are characterized in general by small correlator κ values. We expect that 2DRanking will find further useful applications for large scale gene regulation networks.

D. Networks of game go

The complexity of the well-known game go is such that no computer program has been able to beat a good player, in contrast with chess where world champions have been bested by game simulators. It is partly due to the fact that the total number of possible allowed positions in go is about 10^{171}, compared to e.g. only 10^{50} for chess (Tromp and Farnebäck , 2007).

It has been argued that the complex network analysis can give useful insights for a better understanding of this game. With this aim a network, modeling the game of go, has been defined by a statistical analysis of the data bases of several important historical professional and amateur Japanese go tournaments (Georgeot and Giraud , 2012). In this approach moves/nodes are defined as all possible patterns in 3×3 plaquettes on a go board of 19×19 intersections. Taking into account all possible obvious symmetry operations the number of non-equivalent moves is reduced to $N = 1107$. Moves which are close in space (typically a maximal distance of 4 intersections) are assumed to belong to the same tactical fight generating transitions on the network.

Using the historical data of many games, the transition probabilities between the nodes may be determined leading to a directed network with a finite size Perron-Frobenius operator which can be analyzed by tools of PageRank, CheiRank, complex eigenvalue spectrum, properties of certain selected eigenvectors and also certain other quantities (Georgeot and Giraud , 2012; Kandiah *et al.*, 2014b). The studies are done for plaquettes of different sizes with the corresponding network size changing from $N = 1107$ for plaquettes squares with 3×3 intersections up to maximal $N = 193995$ for diamond-shape plaquettes with 3×3 intersections plus the four at distance two from the center in the four directions left, right, top, down. It is shown that the PageRank leads to a frequency distribution of moves which obeys a Zipf law with exponents close to unity but this exponent may slightly vary if the network is constructed with shorter or longer sequences of successive moves. The important nodes in certain eigenvectors may correspond to certain strategies, such as protecting a stone and eigenvectors are also different between amateur and professional games. It is also found that the different phases of the game go are characterized by a different spectrum of the G matrix. The obtained results show that with the help of the Google matrix analysis it is possible to extract communities of moves which share some common properties.

The authors of these studies (Georgeot and Giraud , 2012; Kandiah *et al.*, 2014b) argue that the Google matrix analysis can find a number of interesting applications in the theory of games and the human decision-making processes.

E. Opinion formation on directed networks

Understanding the nature and origins of mass opinion formation is an outstanding challenge of democratic societies (Zaller, 1999). In the last few years the enormous development of such social networks as LiveJournal, Facebook, Twitter, and VKONTAKTE, with up to hundreds of millions of users, has demonstrated the growing influence of these networks on social and political life. The small-world scale-free structure of the social networks, combined with their rapid communication facilities, leads to a very fast information propagation over networks of electors, consumers, and citizens, making them very active on instantaneous social events. This invokes the need for new theoretical models which would allow one to understand the opinion formation process in modern society in the 21st century.

The important steps in the analysis of opinion formation have been done with the development of various voter models, described in great detail in (Castellano *et al.*, 2009; Krapivsky *et al.*, 2010). This research field became known as sociophysics (Galam , 1986, 2008). Here, following (Kandiah and Shepelyansky , 2012), we analyze the opinion formation process introducing several new aspects which take into account the generic features of social networks. First, we analyze the opinion formation on real directed networks such as WWW of Universities of Cambridge and Oxford (2006), Twitter (2009) and LiveJournal. This allows us to incorporate the correct scale-free network structure instead of unrealistic regular lattice networks, often considered in voter models. Second, we assume that the opinion at a given node is formed by the opinions of its linked neighbors weighted with the PageRank probability of these network nodes. The introduction of such a weight represents the reality of social networks where network nodes are characterized by the PageRank vector which provides a natural ranking of node importance, or elector or society member importance. In a certain sense, the top nodes of PageRank correspond to a political elite of the social network whose opinion influences the opinions of other members of the society (Zaller, 1999). Thus the proposed PageRank opinion formation (PROF) model takes into account the situation in which an opinion of an influential friend from high ranks of the society counts more than an opinion of a friend from a lower society level. We argue that the PageRank probability is the most natural form of ranking of society members. Indeed, the efficiency of PageRank rating had been well demonstrated for various types of scale-free networks.

The PROF model is defined in the following way. In agreement with the standard PageRank algorithm we determine the probability $P(K_i)$ for each node ordered by PageRank index K_i (using $\alpha = 0.85$). In addition, a network node i is characterized by an Ising spin variable σ_i which can take values $+1$ or 1, coded also by red or blue color, respectively. The sign of a node i is determined by its direct neighbors j, which have PageRank probabilities P_j . For that we compute the sum Σ_i over all directly linked neighbors j of node i:

$$\Sigma_i = a \sum_j (P^+{}_{j,\mathrm{in}} - P^-{}_{j,\mathrm{in}}) + b \sum_j (P^+{}_{j,\mathrm{out}} - P^-{}_{j,\mathrm{out}}) \ , \quad a+b=1 \ , \tag{17}$$

where $P_{j,\mathrm{in}}$ and $P_{j,\mathrm{out}}$ denote the PageRank probability P_j of a node j pointing to node i (ingoing link) and a node j to which node i points to (outgoing link), respectively. Here, the two parameters a and b are used to tune the importance of ingoing and outgoing links with the imposed relation $a+b=1$ ($0 \le a, b \le 1$). The values P^+ and P^- correspond to red and blue nodes, and the spin σ_i takes the value 1 or -1, respectively, for $\Sigma_i > 0$ or $\Sigma_i < 0$. In a certain sense we can say that a large value of parameter b corresponds to a conformist society in which an elector i takes an opinion of other electors to which he/she points. In contrast, a large value of a corresponds to a tenacious society in which an elector i takes mainly the opinion of those electors who point to him/her. A standard random number generator is used to create an initial random distribution of spins σ_i on a given network. The time evolution then is determined by the relation (17) applied to each spin one by one. When all N spins are turned following (17) a time unit t is changed to $t \to t+1$. Up to $N_r = 10^4$ random initial generations of spins are used to obtain statistically stable results. We present results for the number of red nodes since other nodes are blue.

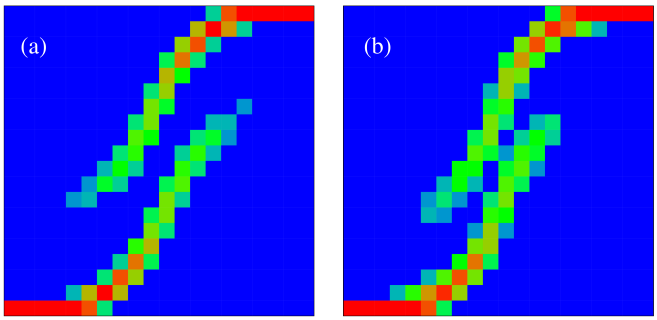

FIG. 57 (Color online) Density plot of probability W_f to find a final red fraction f_f, shown in $y-$axis, in dependence on an initial red fraction f_i, shown in $x-$ axis; data are shown inside the unit square $0 \le f_i, f_f \le 1$. The values of W_f are defined as a relative number of realisations found inside each of 20×20 cells which cover the whole unit square. Here $N_r = 10^4$ realizations of randomly distributed colors are used to obtained W_f values; for each realization the time evolution is followed up the convergence time with up to $t = 20$ iterations; (a) Cambridge network; (b) Oxford network at $a = 0.1$. The probability W_f is proportional to color changing from zero (blue/black) to unity (red/gray). After (Kandiah and Shepelyansky , 2012).

The main part of studies is done for the WWW of Cambridge and Oxford discussed above. We start with a random realization of a given fraction of red nodes $f_i = f(t=0)$ which evolution in time converges to a

steady state with a final fraction of red nodes f_f approximated after time $t_c \approx 10$. However, different initial realisations with the same f_i value evolve to different final fractions f_f clearly showing a bistability phenomenon. To analyze how the final fraction of red nodes f_f depends on its initial fraction f_i, we study the time evolution $f(t)$ for a large number N_r of initial random realizations of colors following it up to the convergence time for each realization. We find that the final red nodes are homogeneously distributed in PageRank index K. Thus there is no specific preference for top society levels for an initial random distribution. The probability distribution W_f of final fractions f_f is shown in Fig. 57 as a function of initial fraction f_i at $a = 0.1$. The results show two main features of the model: a small fraction of red opinion is completely suppressed if $f_i < f_c$ and its larger fraction dominates completely for $f_i > 1 - f_c$; there is a bistability phase for the initial opinion range $f_b \leq f_i \leq 1 - f_b$. Of course, there is a symmetry in respect to exchange of red and blue colors. For the small value $a = 0.1$ we have $f_b \approx f_c$ with $f_c \approx 0.25$. For the larger value $a = 0.9$ we have $f_c \approx 0.35$, $f_b \approx 0.45$ (Kandiah and Shepelyansky , 2012).

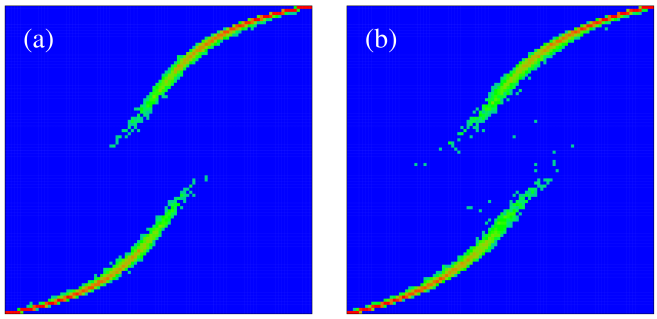

FIG. 58 (Color online) PROF-Sznajd model, option 1: density plot of probability W_f to find a final red fraction f_f, shown in $y-$axis, in dependence on an initial red fraction f_i, shown in $x-$ axis; data are shown inside the unit square $0 \leq f_i, f_f \leq 1$. The values of W_f are defined as a relative number of realizations found inside each of 100×100 cells which cover the whole unit square. Here $N_r = 10^4$ realizations of randomly distributed colors are used to obtained W_f values; for each realization the time evolution is followed up the convergence time with up to $\tau = 10^7$ steps. (a) Cambridge network; (b) Oxford network; here $N_g = 8$. The probability W_f is proportional to color changing from zero (blue/black) to unity (red/gray). After (Kandiah and Shepelyansky , 2012).

Our interpretation of these results is the following. For small values of $a \ll 1$ the opinion of a given society member is determined mainly by the PageRank of neighbors to whom he/she points (outgoing links). The PageRank probability P of nodes to which many nodes point is usually high, since P is proportional to the number of ingoing links. Thus at $a \ll 1$ the society is composed of members who form their opinion by listening to an elite opinion. In such a society its elite with one color opinion can impose this opinion on a large fraction of the society. Indeed, the direct analysis of the case, where the top $N_{top} = 2000$ nodes of PageRank index have the same red color, shows that this 1% of the society elite can impose its opinion to about 50% of the whole society at small a values (conformist society) while at large a values (tenacious society) this fraction drops significantly (see Fig.4 in (Kandiah and Shepelyansky , 2012)). We attribute this to the fact that in Fig. 57 we start with a randomly distributed opinion, since the opinion of the elite has two fractions of two colors this creates a bistable situation when the two fractions of society follow the opinions of this divided elite, which makes the situation bistable on a larger interval of f_i compared to the case of a tenacious society at $a \to 1$. When we replace in (17) P by 1 then the bistability disappears.

However, the detailed understanding of the opinion formation on directed networks still waits it development. Indeed, the results of PROF model for the LiveJournal and Twitted networks show that the bistability in these networks practically disappears. Also e.g. for the Twitter network studied in Sec. X.A, the elite of $N_{top} = 35000$ (about 0.1% of the whole society) can impose its opinion to 80% of the society at small $a < 0.15$ and to about 30% for $a > 0.15$ (Kandiah and Shepelyansky , 2012). It is possible that a large number of links between top PageRank nodes in Twitter creates a stronger tendency to a totalitarian opinion formation comparing to the case of University networks. At the same time the studies of opinion formation with the PROF model on the Ulam networks (Chakhmakhchyan and Shepelyansky , 2013), which have not very large number of links, show practically no bistability in opinion formation. It is expected that a small number of loops is at the origin of such a difference in respect to university networks.

Finally we discuss a more generic version of opinion formation called the PROF-Sznajd model (Kandiah and Shepelyansky , 2012). Indeed, we see that in the PROF model on university network opinions of small groups of red nodes with $f_i < f_c$ are completely suppressed that seems to be not very realistic. In fact, the Sznajd model (Sznajd-Weron and Sznajd , 2000) features the idea of resistant groups of a society and thus incorporates a well-known trade union principle "United we stand, divided we fall". Usually the Sznajd model is studied on regular lattices. Its generalization for directed networks is done on the basis of the notion of group of nodes N_g at each discrete time step τ.

The evolution of group is defined by the following rules:

(a) we pick in the network by random a node i and consider the polarization of $N_g - 1$ highest PageRank nodes pointing to it;

(b) if node i and all other $N_g - 1$ nodes have the same color (same polarization), then these N_g nodes form a group whose effective PageRank value is the sum of all the member values $P_g = \sum_{j=1}^{N_g} P_j$;

(c) consider all the nodes pointing to any member of the group and check all these nodes n directly linked to the group: if an individual node PageRank value P_n is

less than the defined above P_g , the node joins the group by taking the same color (polarization) as the group nodes and increase P_g by the value of P_n; if it is not the case, a node is left unchanged.

The above time step is repeated many times during time τ, counting the number of steps and choosing a random node i on each next step.

The time evolution of this PROF-Sznajd model converges to a steady state approximately after $\tau \approx 10N$ steps. This is compatible with the results obtained for the PROF model. However, the statistical fluctuations in the steady-state regime are present keeping the color distribution only on average. The dependence of the final fraction of red nodes f_f on its initial value f_i is shown by the density plot of probability W_f in Fig. 58 for the university networks. The probability W_f is obtained from many initial random realizations in a similar way to the case of Fig. 57. We see that there is a significant difference compared to the PROF model: now even at small values of f_i we find small but finite values of f_f, while in the PROF model the red color disappears at $f_i < f_c$. This feature is related to the essence of the Sznajd model: here, even small groups can resist against the totalitarian opinion. Other features of Fig. 58 are similar to those found for the PROF model: we again observe bistability of opinion formation. The number of nodes N_g, which form the group, does not significantly affect the distribution W_f (for studied $3 \leq N_g \leq 13$).

The above studies of opinion formation models on scale-free networks show that the society elite, corresponding to the top PageRank nodes, can impose its opinion on a significant fraction of the society. However, for a homogeneous distribution of two opinions, there exists a bistability range of opinions which depends on a conformist parameter characterizing the opinion formation. The proposed PROF-Sznajd model shows that totalitarian opinions can be escaped from by small subcommunities. The enormous development of social networks in the last few years definitely shows that the analysis of opinion formation on such networks requires further investigations.

XV. DISCUSSION

Above we considered many examples of real directed networks where the Google matrix analysis finds useful applications. The examples belong to various sciences varying from WWW, social and Wikipedia networks, software architecture to world trade, games, DNA sequences and Ulam networks. It is clear that the concept of Markov chains and Google matrix represents now the mathematical foundation of directed network analysis.

For Hermitian and unitary matrices there are now many universal concepts, developed in theoretical physics, so that the main properties of such matrices are well understood. Indeed, such characteristics as level spacing statistics, localization and delocalization properties of eigenstates, Anderson transition (Anderson , 1958), quantum chaos features can be now well handled by various theoretical methods (see e.g. (Akemann *et al.*, 2011; Evers and Mirlin , 2008; Guhr *et al.*, 1998; Haake, 2010; Mehta, 2004)). A number of generic models has been developed in this area allowing to understand the main effects via numerical simulations and analytical tools.

In contrast to the above case of Hermitian or unitary matrices, the studies of matrices of Markov chains of directed networks are now only at their initial stage. In this review, on examples of real networks we illustrated certain typical properties of such matrices. Among them there is the fractal Weyl law, which has certain traces in the field of quantum chaotic scattering, but the main part of features are new ones. In fact, the spectral properties of Markov chains had not been investigated on a large scale. We try here to provide an introduction to the properties of such matrices which contain all information about large scale directed networks. The Google matrix is like *The Library of Babel* (Borges, 1962), which contains everything. Unfortunately, we are still not able to find generic Markov matrix models which reproduce the main features of the real networks. Among them there is the possible spectral degeneracy at damping $\alpha = 1$, absence of spectral gap, algebraic decay of eigenvectors. Due to absence of such generic models it is still difficult to capture the main properties of real directed networks and to understand or predict their variations with a change of network parameters. At the moment the main part of real networks have an algebraic decay of PageRank vector with an exponent $\beta \approx 0.5 - 1$. However, certain examples of Ulam networks (see Figs. 13, 14) show that a delocalization of PageRank probability over the whole network can take place. Such a phenomenon looks to be similar to the Anderson transition for electrons in disordered solids. It is clear that if an Anderson delocalization of PageRank would took place, as a result of further developments of the WWW, the search engines based on the PageRank would loose their efficiency since the ranking would become very sensitive to various fluctuations. In a sense the whole world would go blind the day such a delocalization takes place. Due to that a better understanding of the fundamental properties of Google matrices and their dependencies on various system parameters have a high practical significance. We believe that the theoretical research in this direction should be actively continued. In many respects, as *the Library of Babel*, the Google matrix still keeps its secrets to be discovered by researchers from various fields of science. We hope that a further research will allow "*to formulate a general theory of the Library and solve satisfactorily the problem which no conjecture had deciphered: the formless and chaotic nature of almost all the books.*" (Borges, 1962)

XVI. ACKNOWLEDGMENTS

We are grateful to our colleagues M. Abel, A. D. Chepeliankii, Y.-H. Eom, B. Georgeot, O. Giraud, V. Kandiah, O. V. Zhirov for fruitful collaborations on the topics included in this review. We also thank our partners of the EC FET Open project NADINE A. Benczúr, N. Litvak, S. Vigna and colleague A.Kaltenbrunner for illuminating discussions. Our special thanks go to Debora Donato for her insights at our initial stage of this research.

Our research presented here is supported in part by the EC FET Open project "New tools and algorithms for directed network analysis" (NADINE *No* 288956). This work was granted access to the HPC resources of CALMIP (Toulouse) under the allocation 2012-P0110. We also thank the United Nations Statistics Division for provided help and friendly access to the UN COMTRADE database.

References

Abel, M. W., and D. L. Shepelyansky, 2011, Eur. Phys. J. B **84**, 493.

Akemann, G., J. Baik, and Ph. Di Francesco 2011, *The Oxford Handbook of Random Matrix Theory* (Oxford University Press, Oxford).

Albert, R., and A.-L. Barabási, 2000, Phys. Rev. Lett. **85**, 5234.

Albert, R., and A.-L. Barabási, 2002, Rev. Mod. Phys. **74**, 47.

Alon, U., 2014, *U.Alon web site* http://wws.weizmann.ac.il/mcb/UriAlon/ .

Altun, Z.F., L.A. Herndon, C. Crocker, R. Lints, and D.H. Hall (Eds.), 2012, WormAtlas http://www.wormatlas.org.

Anderson, P. W., 1958, Phys. Rev. **109**, 1492.

Aragón, P., D. Laniado, A. Kaltenbrunner, and Y. Volkovich, 2012, Proc. 8th WikiSym2012, ACM, New York **19**, arXiv:1204.3799v2[cs.SI].

Arnoldi, W. E., 1951, Quart. Appl. Math. **9**, 17.

M. Barigozzi, G. Fagiolo, and D. Garlaschelli, 2010, Phys. Rev. E **81**, 046104.

Bascompte, J., P. Jordano, C.J. Melian, and J.M. Olesen, 2003, Proc. Nat. Acad. Sci. USA **100**, 9383.

Bastolla, U., M. A. Pascual-Garcia, A. Ferrera, B. Luque, and J. Bascompte, 2009, Nature (London) **458**, 1018.

Blank, M., G. Keller, and J. Liverani, 2002, Nonlinearity **15**, 1905.

Bohigas, O., M.-J.. Giannoni, and C. Schmit, 1984, Phys. Rev. Lett. **52**, 1.

Borges, J.L., 1962, *The Library of Babel (Ficciones)* (Grove Press, N.Y.).

Brin, S., and L. Page, 1998, Comp. Networks ISDN Syst. **30**, 107.

Brin, M., and G. Stuck 2002, *Introduction to Dynamical Systems* (Cambridge University Press, Cambridge UK).

Bruzda, W., M. Smaczyński, V. Cappellini, H.-J. Sommers, and K. Zyczkowski, 2010, Phys. Rev. E **81**, 066209.

Bullmore, E., and O. Sporns, 2009, Nat. Rev. Neurosci. **10**, 312.

Burgos, E., H. Ceva, R.P.J. Perazzo, M. Devoto, D. Medan, M. Zimmermann,and A.M. Delbue, 2007, J. Theor. Biol. **249**, 307.

Burgos, E., H. Ceva, L. Hernández, R.P.J. Perazzo, M. Devoto, and D. Medan, 2008, Phys. Rev. E **78**, 046113.

Caldarelli, G., 2003, *Scale-free networks* (Oxford Univ. Press, Oxford).

Capocci, A., V. D. P. Servedio, F. Colariori, L. S. Buriol, D. Donato, S. Leonardi, and G. Caldarelli, 2006, Phys. Rev. E **74**, 036116.

Castellano, C., S. Fortunato, and V. Loreto, 2009, Rev. Mod. Phys. **81**, 591.

Chakhmakhchyan, L., and D. L. Shepelyansky, 2013, Phys. Lett. A **377**, 3119.

Chepelianskii, A.D., and D. L. Shepelyansky, 2001, http://www.quantware.ups-tlse.fr/talks-posters/chepelianskii2001.pd.

Chepelianskii, A. D., 2010, arXiv:1003.5455[cs.SE] .

Chirikov, B. V., 1979, Phys. Rep. **52**, 263.

Chirikov, B.V., and D. Shepelyansky, 2008, Scholarpedia **3(3)**, 3550.

Central Intelligence Agency, 2009, *The CIA Wold Factbook 2010* (Skyhorse Publ. Inc.).

Cornfeld, I.P., Fomin, S.V., and Y.G. Sinai 1982, *Ergodic Theory* (Springer, New York).

Craig, B., and G. von Peter 2010, *Interbank tiering and money center bank* (Discussion paper N 12/2010, Deutsche Bundesbank)

De Benedictis, L., and L. Tajoli 2011, The World Economy **34**, 8, 1417–1454.

Dijkstra, E.W., 1982, *Selected Writing on Computing: a Personal Perspective* (Springer-Verlag, New York).

Dimassi, M., and J. Sjöstrand 1999, *Spectral Asymptotics in the Semicalssical Limit* (Cambridge University Press, Cambridge)

Donato, D., L. Laura, S. Leonardi, and S. Millozzi, 2004, Eur. Phys. J. B **38**, 239.

Dorogovtsev, S. N., A. V. Goltsev, and J. F. F. Mendes, 2008, Rev. Mod. Phys. **80**, 1275.

Dorogovtsev, S. 2010, *Lectures on Complex Networks* (Oxford University Press, Oxford).

Ensemble Genome database, 2011, *Ensemble Genome database* http://www.ensembl.org/ .

Eom, Y.-H., and D. L. Shepelyansky, 2013a, PLoS ONE **8(10)**, e74554.

Eom, Y.-H., K. M. Frahm, A. Benczúr, and D. L. Shepelyansky, 2013b, Eur. Phys. J. B **86**, 492.

Eom, Y.-H., P. Aragón, D. Laniado, A. Kaltenbrunner, S. Vigna, and D. L. Shepelyansky, 2014, arXiv:1405.7183 [cs.SI] (submitted to PLoS ONE).

Ermann, L., and D. L. Shepelyansky, 2010a, Phys. Rev. E **81**, 032621.

Ermann, L., and D. L. Shepelyansky, 2010b, Eur. Phys. J. B **75**, 299.

Ermann, L., A. D. Chepelianskii, and D. L. Shepelyansky, 2011a, Eur. Phys. J. B **79**, 115.

Ermann, L., and D. L. Shepelyansky, 2011b, Acta Phys. Polonica A **120(6A)**, A158; http://www.quantware.ups-tlse.fr/QWLIB/tradecheirank/.

Ermann, L., A. D. Chepelianskii, and D. L. Shepelyansky, 2012a, J. Phys. A: Math. Theor. **45**, 275101; http://www.quantware.ups-tlse.fr/QWLIB/dvvadi/.

Ermann, L., and D. L. Shepelyansky, 2012b, Physica D **241**, 514.

Ermann, L., and D. L. Shepelyansky, 2013a, Phys. Lett. A **377**, 250.

Ermann, L., K. M. Frahm, and D. L. Shepelyansky, 2013b, Eur. Phys. J. B **86**, 193.

Evers, F., and A. D. Mirlin, 2008, Rev. Mod. Phys. **80**, 1355.

Felleman, D.J., and D. C. van Essen, 1991, Celeb. Cortex **1**, 1.

FETNADINE database, 2014, *Quantware group*, http://www.quantware.ups-tlse.fr/FETNADINE/datasets.htm

Fogaras, D., 2003, Lect. Not. Comp. Sci. **2877**, 65.

Fortunato, S., 2010, Phys. Rep. **486**, 75.

Frahm, K. M., and D. L. Shepelyansky, 2009, Phys. Rev. E **80**, 016210 .

Frahm, K. M., and D. L. Shepelyansky, 2010, Eur. Phys. J. B **76**, 57.

Frahm, K. M., B. Georgeot, and D. L. Shepelyansky, 2011, J. Phys A: Math. Theor. **44**, 465101.

Frahm, K. M., A. D. Chepelianskii, and D. L. Shepelyansky, 2012a, J. Phys A: Math. Theor. **45**, 405101.

Frahm, K. M., and D. L. Shepelyansky, 2012b, Eur. Phys. J. B **85**, 355.

Frahm, K. M., and D. L. Shepelyansky, 2012c, Phys. Rev. E **85**, 016214.

Frahm, K. M., and D. L. Shepelyansky, 2013, Eur. Phys. J. B **86**, 322.

Frahm, K. M., and D. L. Shepelyansky, 2014a, Eur. Phys. J. B **87**, 93.

Frahm, K. M., Y.-H.. Eom, and D. L. Shepelyansky, 2014b, Phys. Rev. E **89**, 052814.

Franceschet, M., 2011, Communications of the ACM **54(6)**, 92.

Froyland, G., and K. Padberg 2009, Physica D **238**, 1507.

Galam, S., 1986, J. Math. Psych. **30**, 426.

Galam, S., 2008, Int. J. Mod. Phys. C **19**, 409.

Gamow, G. A., 1928, Z. für Phys **51**, 204.

Garlaschelli, D., and M. I. Loffredo 2005, Physica A: Stat. Mech. Appl. **355**, 138.

Garratt, R.J., Mahadeva, L., and K. Svirydzenka 2011, *Mapping systemic risk in the international banking network* (Working paper N 413, Bank of England)

Gaspard, P., 1998, *Chaos, Scattering and Statistical Mechanics* (Cambridge Univ. Press, Cambridge).

Gaspard, P., 2014, Scholarpedia **9(6)**, 9806.

Georgeot, B., O. Giraud, and D. L. Shepelyansky, 2010, Phys. Rev. E **81**, 056109.

Georgeot, B., and O. Giraud, 2012, Eurphys. Lett. **97**, 68002.

Giraud, O., B. Georgeot, and D. L. Shepelyansky, 2005, Phys. Rev. E **72**, 036203.

Giraud, O., B. Georgeot, and D. L. Shepelyansky, 2009, Phys. Rev. E **80**, 026107.

Goldsheid, I.Y., and B.A. Khoruzhenko, 1998, Phys. Rev. Lett. **80**, 2897.

Golub, G. H., and C. Greif, 2006, BIT Num. Math. **46**, 759.

Guhr, T., A. Mueller-Groeling, and H. A. Weidenmueller, 1998, Phys. Rep. **299**, 189.

Haake, F., 2010, *Quantum Signatures of Chaos* (Springer-Verlag, Berlin).

Hart, M.H., 1992, *The 100: ranking of the most influential persons in history* (Citadel Press, N.Y.).

He, J. and M. W. Deem 2010, Phys. Rev. Lett. **105**, 198701.

Hrisitidis, V., H. Hwang, and Y. Papakonstantinou, 2008, ACM Trans. Database Syst. **33**, 1.

Izhikevich, E.M., and G. M. Edelman, 2008, Proc. Nat. Acad. Sci. **105**, 3593.

Kandiah, V., and D. L. Shepelyansky, 2012, Physica A **391**, 5779.

Kandiah, V., and D. L. Shepelyansky, 2013, PLoS ONE **8(5)**, e61519.

Kandiah, V., and D. L. Shepelyansky, 2014a, Phys. Lett. A **378**, 1932.

Kandiah, V., B. Georgeot, and O. Giraud, 2014b, arXiv:1405.6077 [physics.soc-ph] .

Kernighan, B.W., and D.M. Ritchie 1978, *The C Programming Language* (Englewood Cliffs, NJ Prentice Hall).

Kleinberg, J.M., 1999, Jour. of the ACM **46(5)**, 604.

Krapivsky, P.L., S. Redner, and E. Ben-Naim 2010, *A Kinetic View of Statistical Physics* (Cambridge University Press, Cambridge UK).

Krugman, P. R., M. Obstfeld, and M. Melitz, International economics: theory & policy, 2011, Prentic Hall, New Jersey.

Landau, L.D., and E.M. Lifshitz 1989, *Quantum Mechanics* (Nauka, Moscow).

Langville, A.M., and C.D. Meyer 2006, *Google's PageRank and Beyond: The Science of Search Engine Rankings* (Princeton University Press, Princeton).

Li, T.-Y., 1976, J. Approx. Theory **17**, 177.

Lichtenberg, A. J., and M.A. Lieberman 1992, *Regular and Chaotic Dynamics* (Springer, Berlin).

Linux Kernel releases are downloaded from, 2010, http://www.kernel.org/ .

Litvak, N., W. R. W. Scheinhardt, and Y. Volkovich, 2008, Lect. Not. Comp. Sci. (Springer) **4936**, 72.

Lu, W.T., S. Sridhar, and M. Zworski, 2003, Phys. Rev. Lett. **91**, 154101.

Mantegna, R.N., S.V. Buldyrev, A.L. Goldberg, S. Havlin, C.-K. Peng, M. Simons, and H.E. Stanley, 1995, Phys. Rev. E **52**, 2939.

Markov, A. A., 1906, Izvestiya Fiziko- matematicheskogo obschestva pri Kazanskom universitete, 2-ya seriya (in Russian) **15**, 135.

May, R.M., 2001, *Stability and Complexity in Model Ecosystems* (Princeton Univ. Press, New Jersey, USA).

Mehta, M. L., 2004, *Random matrices* (Elsevier-Academic Press, Amsterdam).

Memmott, J., N.M. Waser, and M.V. Price, 2004, Proc. R. Soc. Lond. B **271**, 2605.

Meusel, R., S. Vigna, O. Lehmberg, and C. Bizer, 2014, Proc. WWW'14 Companion, http://dx.doi.org/10.1145/2567948.2576928.

Milo, R., S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, 2002, Science **298**, 824.

Muchnik, L., R. Itzhack, S. Solomon, and Y. Louzoun, 2007, Phys. Rev. E **76**, 016106.

von Neumann, J., 1958, *The Computer and The Brain* (Yale Univ. Press, New Haven CT).

Newman, M. E. J., 2001, Proc. Natl. Acad. Sci. USA **98**, 404.

Newman, M. E. J., 2003, SIAM Review **45**, 167.

Newman, M. E. J., 2010, *Networks: An Introduction* (Oxford University Press, Oxford UK).

Nonnenmacher, S., and M. Zworski, 2007, Commun. Math. Phys. **269**, 311.

Nonnenmacher, S., J. Sjoestrand, and M. Zworski, 2014, Ann. Math. **179**, 179.

Olesen, J.M., J. Bascompte, Y.L. Dupont, and P. Jordano, 2007, Proc. Natl. Acad. Sci. USA **104**, 19891.

Pandurangan, G., P. Raghavan, and E. Upfal, 2005, Internet Math. **3**, 1.

Pantheon MIT project, 2014, *Pantheon MIT project*

http://pantheon.media.mit.edu .
Perra, N., V. Zlatic, A. Chessa, C. Conti, D. Donato, and G. Caldarelli, 2009, Europhys. Lett. **88**, 48002.
Radicchi, F., S. Fortunato, B. Markines, and A. Vespignani, 2009, Phys. Rev. E **80**, 056103.
Redner, S., 1998, Eur. Phys. J. B **4**, 131.
Redner, S., 2005, Phys. Today **58(6)**, 49.
Rezende, E.L., J.E. Lavabre, P.R. Guimaraes, P. Jordano, and J. Bascompte, 2007, Nature (London) **448**, 925.
Rodríguez-Gironés, M.A., and L. Santamaría, 2006, J. Biogeogr. **33**, 924.
Saverda, S., D.B. Stouffer, B. Uzzi, and J. Bascompte, 2011, Nature (London) **478**, 233.
Serra-Capizzano, S., 2005, SIAM J. Matrix Anal. Appl. **27**, 305.
Serrano, M. A., M. Boguna, and A. Vespignani, 2007, J. Econ. Interac. Coor. **2**, 111.
Shanghai ranking, 2010, Academic ranking of world universities http://www.shanghairanking.com/ .
Shen-Orr, A., R. Milo, S. Mangan, and U. Alon, 2002, Nature Genetics **31(1)**, 64.
Shepelyansky, D. L., 2001, Phys. Scripta **T90**, 112.
Shepelyansky, D. L., 2008, Phys. Rev. E **77**, 015202(R).
Shepelyansky, D. L., and O. V. Zhirov, 2010a, Phys. Rev. E **81**, 036213.
Shepelyansky, D. L., and O. V. Zhirov, 2010b, Phys. Lett. A **374**, 3206.
Sjöstrand, J., 1990, Duke Math. J. **60**, 1.
SJR, 2007, SCImago. (2007). SJR SCImago Journal & Country Rank http://www.scimagojr.com .
Skiena, S., and C.B. Ward 2014, *Who's bigger?: where historical figures really rank* (Cambridge University Press, New York); http://www.whoisbigger.com/.
Soramäki, K., M. L. Bech, J. Arnold, R. J. Glass, and W. E. Beyeler, 2005, Physica A **379**, 317.
Song, C., S. Havlin, and H.A. Makse, 2005, Nature (London) **433**, 392.
Sporns, O., 2007, Scholarpedia **2(10)**, 4695.
Stewart, G. W., 2001, *Matrix Algorithms Vol. II: Eigensystems* (SIAM, Philadelphia PA).
Sznajd-Weron, K., and J. Sznajd, 2000, Int. J. Mod. Phys. C **11**, 1157.
Towlson, E.K., P. E. Vértes, S.E. Ahnert, W.R. Schafer, and E.T. Bullmore, 2013, J. Neurosci. **33(15)**, 6380.
Tromp, J., and G. Farnebäck, 2007, Lect. Notes Comp. Sci. (Springer) **4630**, 84.
UK universities, 2011, *Academic Web Link Database* http://cybermetrics.wlv.ac.uk/database/ .
Ulam, S., 1960, *A Collection of Mathematical Problems* (Interscience Tracs in Pure and Applied Mathematics, Interscience, New York).
UN COMTRADE, 2011, *United Nations Commodity Trade Statistics Database* http://comtrade.un.org/db/ .
Vázquez, D.P., and M. A. Aizen, 2004, Ecology **85**, 1251.
Vigna, S., 2013, arXiv:0912.0238v13[cs.IR] .
Watts, D. J., and S. H. Strogatz, 1998, Nature (London) **393**, 440.
Weyl, H., 1912, Math. Ann. **141**, 441.
Wikipedia top 100 article, 2014, *Top 100 historical figures of Wikipedia* http://www.wikipedia.org/en
Zaller, J.R., 1999, *The Nature and Origins of Mass Opinion* (Cambridge University Press, Cambridge UK).
Zhirov, A. O., O. V. Zhirov, and D. L. Shepelyansky, 2010, Eur. Phys. J. B **77**, 523; http://www.quantware.ups-tlse.fr/QWLIB/2drankwikipedia/ .
Zipf, G. K., 1949, *Human Behavior and the Principle of Least Effort* (Addison-Wesley, Boston).
Zlatic, V., M. Bozicevic, H. Stefancic, and M. Domazet, 2006, Phys. Rev. E **74**, 016115.
Zuo, X.-N., R. Ehmke, M. Mennes, D. Imperati, F.X. Castellanos, O. Sporns, and M.P. Milham, 2012, Cereb. Cortex **22**, 1862.
Zworski, M., 1999, Not. Am. Math. Soc. **46**, 319.
Zyczkowski, K., M. Kus, W. Slomczynski, and H.-J. Sommers, 2003, J. Phys. A: Math. Gen. **36**, 3425.

www.ann-phys.org

adp

annalen der physik

WILEY-VCH

Wiley Online Library
Early View publication on www.wileyonlinelibrary.com
(issue and page numbers not yet assigned;
citable using Digital Object Identifier – **DOI**)

Ann. Phys. (Berlin), 1–10 (2015) / DOI 10.1002/andp.201500110

annalen der physik

Original Paper

Anderson transition for Google matrix eigenstates

O. V. Zhirov[1,2] *and D. L. Shepelyansky*[3,*]

Received 2 February 2015, revised 20 February 2015, accepted 20 February 2015
Published online 16 March 2015

We introduce a number of random matrix models describing the Google matrix G of directed networks. The properties of their spectra and eigenstates are analyzed by numerical matrix diagonalization. We show that for certain models it is possible to have an algebraic decay of PageRank vector with the exponent similar to real directed networks. At the same time the spectrum has no spectral gap and a broad distribution of eigenvalues in the complex plain. The eigenstates of G are characterized by the Anderson transition from localized to delocalized states and a mobility edge curve in the complex plane of eigenvalues.

1 Introduction

The phenomenon of Anderson localization [1] appears in a variety of quantum physical systems including electron transport in disordered solids and waves in random media (see e.g. [2, 3]). It is usually analyzed in the frame of Hermitian or unitary matrices. Recently, the localization properties of nonunitary complex matrices has been analyzed for Euclidean matrices [4] in relation to light and wave localization [5].

In this work we analyze the possibilities of Anderson like localization and delocalization for the matrices belonging to the class of Markov chains and Google matrix G [6, 7]. Such matrices have real nonnegative elements with the sum of elements in each column being equal to unity. For a directed network one first defines an adjacency matrix A_{ij} which has element 1 if a node j have a link pointing to node i and zero otherwise. The columns with only zero elements (*dangling nodes*) are replaced by columns with $1/N$ where N is the matrix size. The elements of other columns are renormalized in such a way that their sum becomes equal to unity ($\sum_i S_{ij} = 1$, $S_{ij} = A_{ij}/\sum_i A_{ij}$). Thus we obtain the matrix S_{ij} of Markov transitions. Then the Google matrix G of the network takes the form [6, 7]:

$$G_{ij} = \alpha S_{ij} + (1-\alpha)/N. \qquad (1)$$

Here, the damping factor α is taken in the range $0 < \alpha \leq 1$. In the context of the World Wide Web (WWW) the term $(1-\alpha)$ describes for a random surfer a probability to jump on any node of the network. The above construction of G has been proposed by Brin and Page [6] to describe the structure of the World Wide Web (WWW). For the WWW it is assumed that the Google search engine uses $\alpha \approx 0.85$ [7]. We can also consider a generalized case of weighted Markov transitions S_{ij} corresponding to real positive elements of A_{ij} like happens for the world trade network (see e.g. [8]).

The matrix G belongs to the class of Perron-Frobenius operators, its largest eigenvalue is $\lambda = 1$ and other eigenvalues have $|\lambda| \leq \alpha$ [7, 9]. The right eigenvector at $\lambda = 1$, which is called the PageRank ($GP = P$), has real nonnegative elements $P(i)$ and gives a probability $P(i)$ to find a random surfer at site i. It is possible to rank all nodes in a decreasing order of PageRank probability $P(K(i))$ so that the PageRank index $K(i)$ counts all N nodes i according their ranking, placing the most popular nodes at the top values $K = 1, 2, 3....$ Usually for many real directed networks the distributions of number of ingoing and outgoing links are described by a power law (see e.g. [10]), generating an average approximately algebraic decay of PageRank probability $P(K) \propto 1/K^{\beta}$ with $\beta \approx 0.9$. Some examples of directed networks can be found in [11].

It is important to note that matrices of Google class practically have not been studied in physical systems even if they naturally appear in the frame of Ulam networks generated by the Ulam method for dynamical maps in a coarse-grained phase space (see e.g. [12–14]).

Therefore, it is interesting to see if the phenomena of Anderson localization and Anderson delocalization transition can appear for Google matrices. Certain indications on a possible Anderson transition for the Ulam networks, built from dissipative maps, have been reported

* Corresponding author E-mail: dima@irsamc.ups-tlse.fr
[1] Budker Institute of Nuclear Physics, 630090 Novosibirsk, Russia
[2] Novosibirsk State University, 630090 Novosibirsk, Russia
[3] Laboratoire de Physique Théorique du CNRS (IRSAMC), Université de Toulouse, UPS, F-31062 Toulouse, France

Original Paper

in [12] with more detailed discussions presented in [11]. Thus, it would be useful to find random matrix models which are able to reproduce typical properties of spectrum and PageRank decay in real directed networks. However, the results presented in [15] show that the full matrix G with random matrix elements have an unrealistic spectrum and hence other random matrix models of G should be developed. The models discussed in [16] give certain indications of delocalization of eigenstates of G but the spectrum of G in these models has a large gap and is very far from the spectra of real directed networks. With this aim we describe below a number of random Google matrix models and analyze the properties of their spectra and eigenstates. We use certain spectral properties of small size *orthostochastic* matrices with $N = 3, 4$ established in [17].

2 Random matrix models of G

We start from a description of various random matrix models of Google matrix G presenting the results of their spectral properties in next Section.

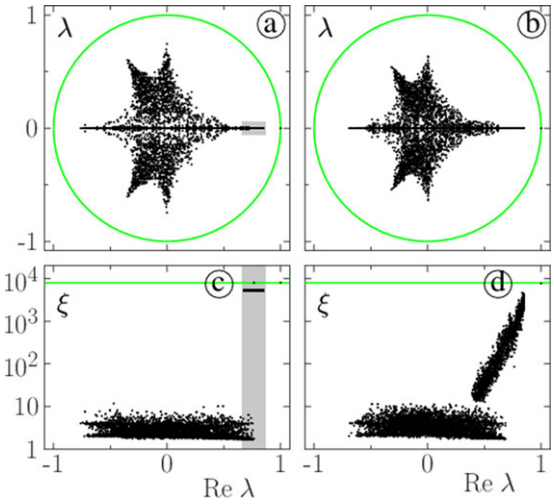

Figure 1 Google matrix eigenvalues λ (a,b), and IPR ξ of eigenvectors as a function of Reλ (c,d). Panels show data for RMZ3 model (Sec. 2.1) at fixed amplitudes $\varepsilon_i = 0.5$ (a,c) and for random amplitudes $0.15 \leq \varepsilon_i \leq 0.3$ (b,d). The green circle shows $|\lambda| = 1$ (a,b); the green horizontal line shows maximal possible $\xi = N$(c,d); the gray band in (a,c) highlights specific states (see text). Here the total number of nodes is N=8000.

2.1 Model RMZ3: random three-diagonal blocks

Following [17] we consider *orthostochastic* matrix blocks B_{ij} of size $M \times M = 4 \times 4$. The *orthostochastic* property means that $B_{ij} = {O_{ij}}^2$, where an orthogonal matrix O has random matrix elements obtained via random rotations. Since O is an orthogonal matrix the matrix B is bistochastic with $\sum_i B_{ij} = \sum_j B_{ij} = 1$ [17]. The main reason to use such blocks B is a similarity of complex spectrum of random matrix ensemble of B with the spectrum of G of university networks of Cambridge and Oxford, as discussed in [11]. The size 4×4 can be considered as preferential random links between a group of 4 friends. However, a weak point of the random ensemble of B [17] is a small matrix size $N = 4$, while for the above universities we have $N \approx 2 \times 10^5$.

To go to large values of N in matrix S_{ij} we construct the **Random Matrix model** $Z3$ **(RMZ3)** as follows: we place blocks B of size $M = 4$ on the main diagonal with weights $(1 - \varepsilon_i)$ and on two adjacent upper and lower diagonals with weight $\varepsilon_i/2$, where ε_i $(i = 1, \ldots, N/M)$ are random numbers uniformly distributed in some interval $(\varepsilon_{min}, \varepsilon_{max})$; each block represents a random realization of B; then the matrix G of total size N is built from S via the equation (1). Here we consider two cases with a constant $\varepsilon_i = 0.5$ and the interval range $0.15 \leq \varepsilon_i \leq 0.3$ (see Fig. 1). Obviously, by construction the final matrix belongs to the Google matrix class. We use notations S_Z and G_Z for the matrices S and G of this model.

2.2 Model RMZ3S: RMZ3 with shortcuts

The model **RMZ3S** is obtained from RMZ3 by adding shortcut links between blocks B in the upper triangle of the whole matrix S, the blocks of shortcut links are placed randomly in this part of S. The amplitude of transitions from one block to another block (outside of three-diagonal blocks of RMZ3) is taken at some fixed value ε_s. The shortcut blocks are randomly and uniformly distributed over the upper triangle of the whole matrix. *After* adding the shortcut blocks the columns affected by shortcut blocks are renormalized to unity. In this way the obtained matrix S again belongs to the Google matrix class. The blocks of shortcuts are placed randomly in the upper triangle of matrix S, their number N_s is determined by the parameter $\delta = 4N_s/(3N)$. In fact δ gives the ratio of shortcut blocks to the number of blocks $3N/4$ in the model RMZ3. Again each block B in the main three-diagonal part of RMZ3 and blocks at shortcut positions are taken as random and independent realizations for each block. We note that the shortcuts between single nodes have been used for studies of quantum chaos and Anderson transition in the small world Anderson model

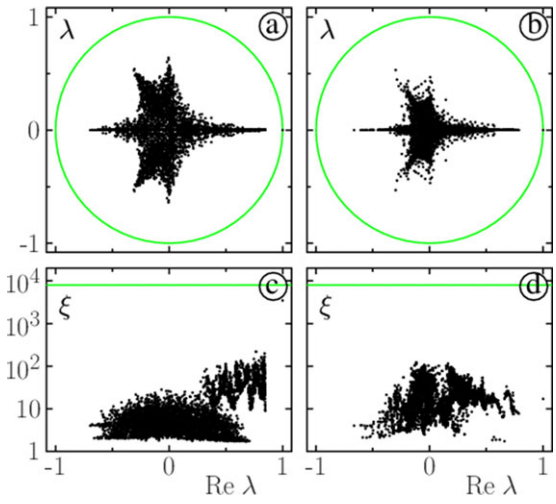

Figure 2 Same as in Fig. 1 for RMZ3S model (Sec. 2.2) with $0.15 \leq \varepsilon_i \leq 0.3$, shortcut amplitude $\varepsilon_s = 0.3$, $\delta = 0.1$ (a,c) and $\delta = 1$ (b,d). Here $N = 8000$.

(see [11, 18, 19]). The results for RMZ3S model are shown in Figs. 2, 3.

2.3 Model RMZ3F: RMZ3 plus triangular matrix

The results obtained in [15, 20] show that a triangular matrix of Google matrix class has a tendency to have a realistic PageRank probability decay with $P \propto 1/K$ and have some eigenvalues of finite amplitudes $|\lambda|$. Due to these indications we construct a matrix S_F in the following way: N_u random numbers f_i from the interval $(0, 1)$ are placed on random positions of the upper triangle of matrix of size N, then all columns are renormalized to unity and columns with all zero elements are replaced by columns with all elements $1/N$. Then we construct the matrix G of the **model RMZF** as:

$$S_{ZF} = (1-\mu)S_Z + \mu S_F\,,\;\; G_{ZF} = \alpha S_{ZF} + (1-\alpha)/N\,. \qquad (2)$$

Here μ determines a measure of contribution of S_F with $0 < \mu < 1$. The number of nonzero random elements N_u is given by parameter $\delta = N_u/(12N)$. The results for the RMZF model are shown in Figs. 3, 4, 5.

2.4 Anderson models AD2 and AD3 for G matrix

We use the usual Anderson model [1, 3] with diagonal disorder terms W_i and transitions V to nearby sites on a

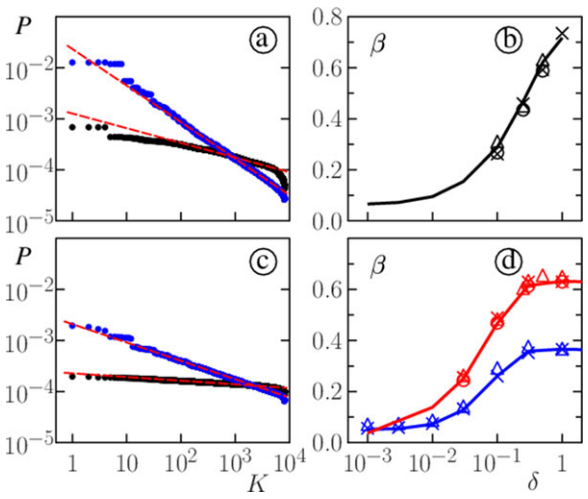

Figure 3 Dependence of $P(K)$ for models RMZ3S (Sec. 2.2) in panels (a,b) and RMZ3F (Sec. 2.3) in panels (c,d); here $0.15 \leq \varepsilon_i \leq 0.3$. In panel (a) we have $\delta = 0.1$ (black symbols) and $\delta = 1$ (blue symbols) at $N = 8000$; the fitted algebraic dependence is shown by straight dashed lines with parameters: $a = -6.67$, $\beta = 0.288$ at $\delta = 0.1$ and $a = -3.76$, $\beta = 0.71$ at $\delta = 1$; panel (b) shows the dependence of β on δ with the full curve for $N = 8000$ and triangles, crosses and circles for $N = 2000, 4000, 16000$ respectively; the amplitude of shortcut elements is $\varepsilon_s = 0.3$. In panel (c) we have $\delta = 0.01$ (black symbols) and $\delta = 3$ (blue symbols) at $\mu = 0.1$ and $N = 8000$; the fitted algebraic dependence is shown by straight dashed lines with parameters: $a = -8.39$, $\beta = 0.072$ at $\delta = 0.01$ and $a = -6.17$, $\beta = 0.36$ at $\delta = 3$; panel (d) shows the dependence of β on δ for $\mu = 0.1$ (blue) and $\mu = 0.3$ (red) with the full curves for $N = 8000$ and triangles, crosses and circles for $N = 2000, 4000, 16000$ respectively.

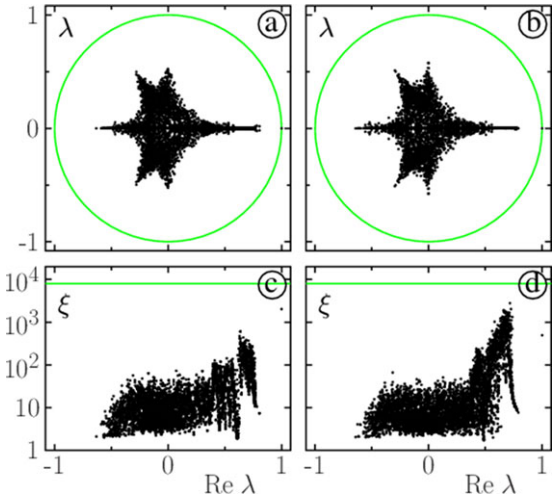

Figure 4 Spectrum (a,b) and IPR ξ dependence of $\mathrm{Re}\lambda$ for the model RMZ3F (Sec. 2.3) at $\delta = 0.1$ (a,c) and $\delta = 3$ (b,d); here $0.15 \leq \varepsilon_i \leq 0.3$, $\mu = 0.1$, $N = 8000$; circle and horizontal lines are as in Fig. 1.

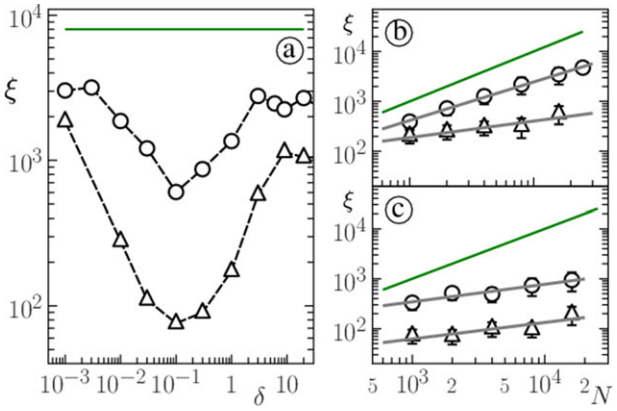

Figure 5 Panel (a) shows dependence of maximal IPR ξ (for states with $|\lambda| < 1$) on parameter δ for the model RMZ3F (Sec. 2.3) at $N = 8000$. Dependence of maximal IPR ξ on N is shown in panels (b) at $\delta = 3$ and (c) at $\delta = 0.1$; error bars show statistical error, if it is larger than symbol size, obtained from N_r disorder realizations. We use $N_r = 11$ at $N = 2000$, $N_r = 8$ at $N = 4000$, $N_r = 4$ at $N = 8000$, $N_r = 3$ at $N = 16000$. In all panels $\mu = 0.1$ (circles) and $\mu = 0.3$ (triangles), $0.15 \leq \varepsilon_i \leq 0.3$; the straight green lines show dependence $\xi = N$; the straight gray lines in (b,c) show the fitted dependence (see text).

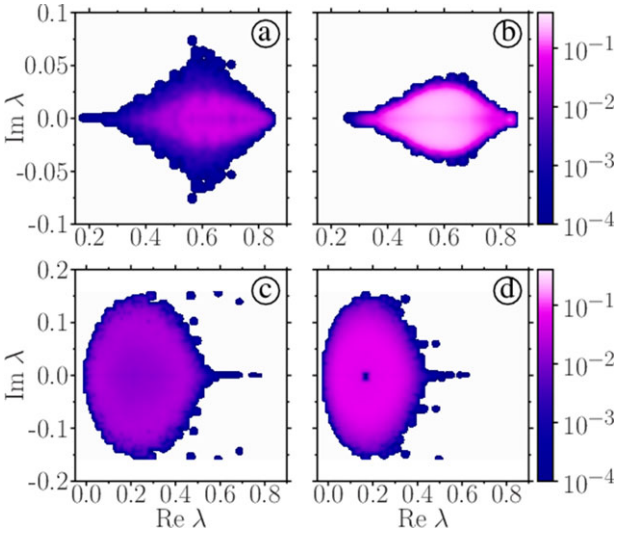

Figure 6 Distribution of IPR ξ_i on $\lambda-$plane for the Anderson type models AD2 at $d = 2$ (a) and AD3 at $d = 3$ (b) (Sec. 2.4) and the Anderson type models with shortcuts AD2S (c) and AD3S (d) at $\delta = 2$ (Sec. 2.5). Here $\varepsilon_{max} = 0.6$; $N = 130^2 = 16900$ for (a,c); $N = 25^3 = 15625$ for (b,d) and $\varepsilon_{max} = 0.6$, $\varepsilon_s = \varepsilon_{max}/2 = 0.3$ for (c,d). Color bars show the ratio ξ_i/N (IPR values are averaged inside cells of coarse-grained lattice 60×60).

lattice in dimension d:

$$W_{\mathbf{i}}\psi_{\mathbf{i}} + V\psi_{\mathbf{i+1}} + V\psi_{\mathbf{i-1}} = \lambda\psi_{\mathbf{i}} \,, \qquad (3)$$

where indexes in bold are vectors in d-dimensional space. On the basis of (3) we construct the matrices S and G.

Thus we consider the dimensions $d = 2, 3$ corresponding to square and cubic lattices. The matrix S is constructed as follows: each transition matrix element, corresponding to V terms, in the Anderson model in dimension d (3) is replaced by a random number ε_i uniformly distributed in the interval $[0, \varepsilon_{max}/2d]$, the diagonal element $W_{\mathbf{i}}$ is replaced by unity minus the sum of all ε_i over $2d$ nearby sites $(1 - \sum_{i=1}^{2d} \varepsilon_i)$. The asymmetric matrix S constructed in this way belongs to the Google matrix class. Thus we obtain the matrices S_{AD2}, G_{AD2} for the **model AD2** and S_{AD3}, G_{AD3} for the **model AD3** for $d = 2$ and 3 respectively. The results for these models are presented in Figs. 6, 7, 9.

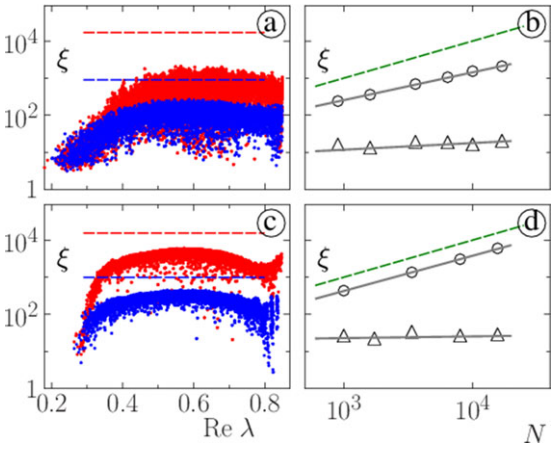

Figure 7 Dependence of ξ on $\mathrm{Re}\lambda$ (a,c) and ξ on N (b,d) for the models AD2 (a,b) and AD3 (b,d) (see Sec. 2.4). For AD2: panel (a) is for $N = 900$ (blue, $N_r = 10$ realizations) and $N = 16900$ (red, $N_r = 1$); panel (b) shows dependence $\xi(N)$ with fits $\xi \propto N^{\nu}$ for eigenstates at the spectrum edge with $\mathrm{Re}\lambda = 0.23 - 0.25$ (triangles, $\nu = 0.18$) and for maximal ξ (circles, $\nu = 0.75$). For AD3: panel (c) is for $N = 1000$ (blue, $N_r = 10$ realizations) and $N = 15625$ (red, $N_r = 1$); panel (d) shows dependence $\xi(N)$ for eigenstates at the spectrum edge with $\mathrm{Re}\lambda = 0.23 - 0.25$ (triangles, $\nu = 0.05$) for maximal ξ (circles, $\nu = 0.95$). Here $\varepsilon_{max} = 0.6$. The fits are shown by gray lines, green (b,d) and blue, red (c,d) dashed lines show dependence $\xi = N$. For panels (b,d) the number of realizations changes from $N_r = 10$ to 3 when N changes from minimal to maximal value.

2.5 Anderson models AD2S and AD3S with shortcuts

By adding shortcut links between pairs of nodes randomly distributed in the upper triangle of matrix S we obtain **models AD2S and AD3S** respectively from models AD2 and AD3. The number of shortcut elements in

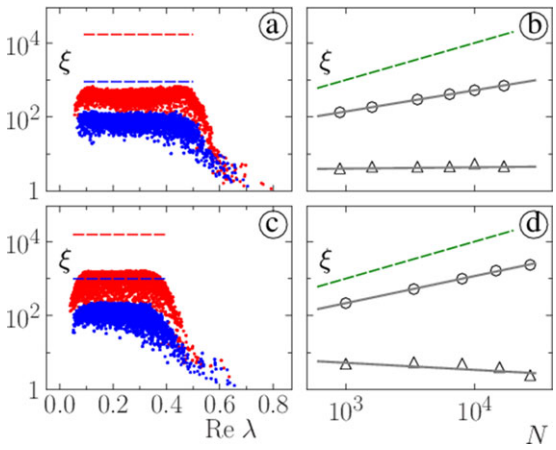

Figure 8 Same as in Fig. 7 but for the models AD2S (a,b) and AD3S (c,d) (see Sec. 2.5) at $\delta = 2$; all parameters are as in Fig. 7. The fits give: (b) $\nu = 0.04$ at the spectrum edge around $\mathrm{Re}\lambda \approx 0.6$ (triangles), $\nu = 0.57$ for maximal ξ (circles); (d) $\nu = -0.19$ at the spectrum edge around $\mathrm{Re}\lambda \approx 0.6$ (triangles), $\nu = 0.73$ for maximal ξ (circles). Here $\varepsilon_{max} = 0.6$, $\varepsilon_s = \varepsilon_{max}/2 = 0.3$. For panel (b) [(d)] the number of realizations changes from $N_r = 10$ to 3 [1] when N changes from 900 to 16900 [27000].

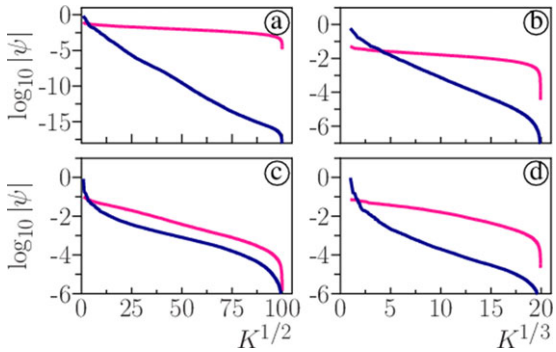

Figure 9 Dependence of eigenvector amplitudes $|\psi|$ on their rank index K for models AD2 (a), AD3 (b) from (Sec. 2.4) and AD2S (c), AD3S (d) from (Sec. 2.5). Here $\delta = 0$ for (a,b) and $\delta = 2$ for (c,d); $N = 10^4$ for (a,c) and $N = 20^3$ for (b,d). We use $\varepsilon_{max} = 0.6$ and $\varepsilon_s = 0.3$ in (c,d). Data show maximally delocalized (maximal ξ corresponding to PageRank, magenta upper curve) and maximally localized (smallest ξ, blue bottom curve) eigenstates.

S is taken to be $N_s = 2dN\delta$, their amplitude is taken as $0 \le \varepsilon_i \le \varepsilon_s = \varepsilon_{max}/2$, after adding shortcuts the columns with shortcut elements are renormalized to unity. Thus the sum of elements in each column is equal to unity and S belongs to the Google matrix class. We note the matrices of these models as S_{AD2S}, G_{AD2S}, S_{AD3S}, G_{AD3S} respectively for $d = 2, 3$. The results for these models are presented in Figs. 8, 9, 10.

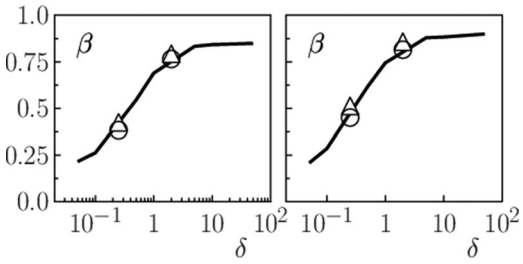

Figure 10 Dependence of the PageRank exponent β on the parameter δ for the models AD2S (left panel) and AD3S (right panel). Left panel: the solid curve shows data for $N = 80^2$, with triangles for $N = 40^2$ and circles for $N = 130^2$. Right panel: the solid curve shows data for $N = 20^3$, with triangles for $N = 10^3$ and circles for $N = 25^3$. Here $\varepsilon_{max} = 0.6$ and $\varepsilon_s = 0.3$.

2.6 Anderson models AD2Z and AD2ZS with blocks and block shortcuts

By replacing matrix elements in the model AD2 by blocks B of size 4×4 (see Sec. 2.1) we obtain the **model AD2Z**. In a similar way for the model AD2S we obtain the **model AD2ZS** with block shortcuts. In this case we restrict our studies only for dimension $d = 2$ since the matrix size becomes too large for $d = 3$. Amplitudes ε_{max} and ε_s are defined as for the models AD2 and AD2S. Since the transitions are now given by blocks then the parameter δ is now defined as $N_s = 2d(N/4)\delta$ with $d = 2$. The results for models AD2Z and AD2ZS are presented in Figs. 11, 12.

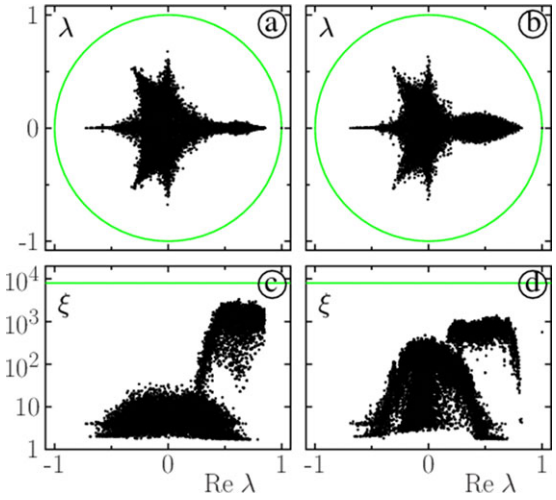

Figure 11 Spectrum λ (a,b) and IPR ξ vs. $\mathrm{Re}\lambda$ (c,d) for the models AD2Z (a,c) and AD2ZS at $\delta = 0.25$ (b.d) from Sec. 2.6. Here $N = 4 \times 70^2 = 19600$, $\varepsilon_{max} = 0.6$ and $\varepsilon_s = 0.3$ in (c,d).

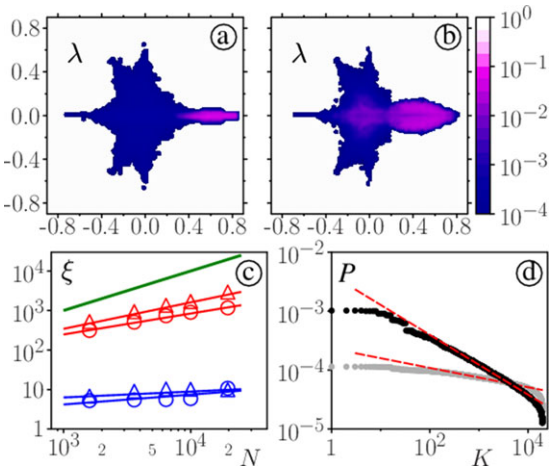

Figure 12 Top panels show distribution of IPR ξ values on $\lambda-$plane for models AD2Z (a) and AD2ZS at $\delta = 0.25$ (b) of Sec. 2.6 with parameters of Fig. 11; color bar gives the ratio ξ/N obtained from cells as in Fig. 6. Panel (c): dependence of ξ on N for AD2Z with triangles for states with λ located in the delocalized domain $\mathrm{Re}\lambda \in (0.3, 0.85)$ (red triangles, fit gives $\nu = 0.67$) and in the localized domain $\mathrm{Re}\lambda < -0.5$ (blue triangles, $\nu = 0.15$); for AD2ZS at $\delta = 0.25$ with circles for states with λ located in the delocalized domain $\mathrm{Re}\lambda \in (0.2, 0.85)$ (red circles, $\nu = 0.53$) and in the quasi-localized domain $\mathrm{Re}\lambda < -0.5$ (blue circles, $\nu = 0.25$); fits are shown by lines, green line shows $\xi = N$. Panel (d): dependence of PageRank probability P on PageRank index K for models AD2Z (gray symbols) and AD2ZS at $\delta = 0.25$ (black symbols); the fits for the range $K \in (100, 6000)$ are shown by dashed lines with $\beta = 0.16$ (AD2Z) and $\beta = 0.51$ (AD2ZS) for the parameters of panels (a,b).

3 Spectral properties of G matrix models

We use exact numerical diagonalization for analysis of spectrum and eigenstates of models of Sec. 2. The matrix size N is changed from a minimal $N = 900$ up to maximal $N = 27000$. For the description of the decay of PageRank probability we use a fit $\ln P = a - \beta \ln K$ which gives us the PageRank exponent of algebraic β. In all simulations we use $\alpha = 0.85$. The right eigenstates $\psi_i(j)$ of G are determined by

$$\sum_{j'=1}^{N} G_{jj'} \psi_i(j') = \lambda_i \psi_i(j) \ . \qquad (4)$$

We characterize $\psi_i(j)$ by the Inverse Participation Ratio (IPR) $\xi_i = (\sum_j |\psi_i(j)|^2)^2 / \sum_j |\psi_i(j)|^4$. This quantity is broadly used in the studies of Anderson localization [3] and determines the number of sites effectively populated by an eigenstate. The value of ξ is independent of normalization. We use normalization $\sum_i P(i) = 1$ for the PageRank eigenstate at $\lambda = 1$. For each eigenvector $\psi_i(j)$ we can order all nodes in a monotonically decreasing order of $|\psi_i(j)|$ thus obtaining the local rank index K for a given $\psi_i(j)$. Such a ranking was used in [21, 22].

We note that in the following Figs. when we show the dependence of ξ on $Re\lambda$ then all λ values are shown; we use various number of disorder realizations which is given in figure captions or equal to one if not directly stated (but we checked that the results are not sensitive to a change of disorder realization).

In the following we present analysis of right eigenstates of G which produce influence on the PageRank vector. The properties of left eigenvectors reserve a separate analysis (e.g. the left eigenvector at $\lambda = 1$ has constant equal elements since the sum of each column elements is equal to unity).

Below we describe the results for models of Sec. 2.

3.1 Results for RMZ3 model

For the model RMZ3 at $\varepsilon_i = const = 0.5$ the spectrum is shown in Fig. 1a. We see that it has a form of 6−rays star typical for the directed networks studied in [11, 21, 22]. The size of the star is slightly reduced since all $\lambda_i(\alpha) \to \alpha\lambda_i(\alpha = 1)$ for $\alpha < 1$, except $\lambda = 1$ [7]. There is also additional reduction of $|\lambda_i|$ due to finite coupling terms $\varepsilon_i > 0$ but this reduction is moderate and the spectrum of G_Z is close to the spectrum of independent 4×4 blocks found in [17]. Thus RMZ3 model captures a part of real properties of directed networks.

An interesting property of eigenstates becomes visible from the dependence of ξ on $\mathrm{Re}\lambda$ shown in Fig. 1c at $\varepsilon_i = 0.5$. Many eigenstates have relatively small $\xi < 10$ which remain bounded with the increase of N up to $N = 16000$ (data not shown). However, there is a group of states (gray band) with $\xi \sim N$ growing linearly with N (data not shown). These are delocalized states. Their origin becomes clear from the following consideration. We can use the anzats in which the elements of $\psi(j)$ are constant inside a given block B_m with a values φ_m. Then Eq.(4) takes the form

$$(1-\varepsilon)\varphi_m + \varepsilon(\varphi_{m+1} + \varphi_{m-1})/2 = \lambda\varphi_m \ , \qquad (5)$$

since the matrix G is bistochastic with sum of elements in rows being unity since $\varepsilon_i = const$. The spectrum λ in (5) is real. Thus we obtain in (5) the Bloch equation with plane wave delocalized solutions well known for crystals [2, 3]. These solutions belong to the gray band part of the spectrum in Fig. 1a. Another part of the spectrum

corresponds to such $\psi(j)$ which have different values on a scale of one block B.

The case with different ε_m (e.g. $0.15 \leq \varepsilon_m \leq 0.3$ in Fig. 1b,d) we can use the same anzats for the left vector φ_m that leads to the eigenvalue equation:

$$(1-\varepsilon_m)\varphi_m + \varepsilon_m(\varphi_{m+1}+\varphi_{m-1})/2 = \lambda\varphi_m. \quad (6)$$

Such a problem corresponds to the case of off-diagonal disorder in the $1d$ Anderson model where the localization length, and hence IPR, is diverging at the center of the band [2, 3]. The spectrum λ in (6) is real. A similar problem is known as the Sinai walk [23] where transition probabilities on a Markov chain are fluctuating. This model has been studied extensively (see e.g. [24] and Refs. therein).

The spectrum λ in (6), corresponding to this anzats, is the same for the right eigenvectors [7]. The right eigenvectors are different from the left ones but have a similar structure on average. The IPR values, shown in Fig. 1d are significantly reduced, comparing to the case $\varepsilon_m = const$, except those with λ close to unity. When N is increasing we find that IPR is growing only for $\lambda \to 1$ while for $|\lambda| < 1$ IPR values remains finite. This corresponds to the known results for the Anderson model with off-diagonal disorder. Other eigenstates for which ψ_i is not constant inside B blocks correspond to the eigenstates with rather small IPR values $\xi \sim 10$.

Even if the spectrum and eigenstates have interesting properties in the two above cases of model RMZ3 there is a weak point here: the PageRank probability P in these cases is flat being practically independent of K and $\xi \sim N$. Thus the situation is very different from the real directed networks with $\beta \approx 1$ (see e.g. [7, 11]). This happens due to a space homogeneous structure of the matrix G (a part of fluctuations) and thus there is no leading node with a large number of links. Due to that we try to introduce shortcut links as described in the next Sec. 3.2.

3.2 Properties of RMZ3S model

The spectrum and IPR dependence for RMZ3S model with shortcuts are shown in Fig. 2 for two typical values of parameter δ. We see that at small values of δ (e.g. $\delta = 0.1$) the spectrum structure is practically the same as for RMZ3 model. However, for larger values ($\delta = 1$) the size of the spectrum star is decreasing. The values of IPR are significantly reduced at finite values of δ and our data show that the maximal ξ values remain less than $\xi = 200$ even for the largest size $N = 16000$ for $0.1 \leq \delta \leq 1$ for all $|\lambda| < 1$ (data not shown). Thus in this model all eigenstates remain localized.

Even if all states are localized the decay of PageRank is more close to the case of real directed networks. Indeed, the data of Fig. 3a,b show that $P(K)$ have approximately algebraic decay with PageRank index. The fit allows to determine the PageRank exponent β which is small at $\delta \sim 0.1$ and is growing with increase of δ reaching values $\beta \approx 0.75$ at $\delta = 1$. It is important to note that β is independent of N at large N values. Thus the homogeneous random elements in the upper triangle of S matrix allow to obtain β close to unity at large δ. Indeed, in the limit of rather large δ we come to the case of triangular matrix S studied in [15] (and also in [20]) where one obtains an approximate decay $P \propto 1/K$. Indeed, at large δ a sum of elements in a row of G drops approximately as $1/K$ (where K is a row index) leading to $P \propto 1/K$. Indeed, we can say that $P(K) \approx \sum_j G_{Kj} e_j \sim 1/K$, where $e_j = 1/N$ is a homogeneous initial vector, considering this as one iteration of the PageRank algorithm [7]. We note that for the PageRank vector we have $\xi \sim N$ for $\beta < \beta_c = 1/4$.

Thus the model RMZ3S has a reasonable spectrum structure and an algebraic PageRank probability decay. But all eigenstates with $|\lambda| < 1$ remain localized. Thus we go to the analysis of RMZ3F model.

3.3 Results for RMZ3F model

The spectrum and IPR values for the RMZ3F model are presented in Fig. 4. We see that the star spectrum structure is preserved but IPR values are increased in a vicinity $\mathrm{Re}\lambda \approx \alpha$. The examples of $P(K)$ and $\beta(\delta)$ dependencies are shown in Fig. 3c,d. It is important to note that β is independent of N at large N values. Qualitatively, the situation is similar to the model RMZ3S but the effect of δ on localization properties of ξ is more complicated.

Indeed, it is well seen in Fig. 5a that the maximal IPR values (excluding PageRank vectors) are at first reduced with an increase of δ from 10^{-3} up to 0.1 but they are increased when δ goes from 0.1 to 10. The dependence of maximal ξ on N at $\delta = 0.1; 3$ is shown in Fig. 5b,c for $\mu = 0.1; 0.3$. We fit this dependence by a power law $\xi \propto N^{\nu}$ and obtain for $\mu = 0.1$: $\nu = 0.352$ (at $\delta = 0.1$) and $\nu = 0.770$ (at $\delta = 3$); for $\mu = 0.3$: $\nu = 0.33$ (at $\delta = 0.1$ and 3). These results show that there are certain states (except PageRank) that become delocalized in the limit of large matrix size. In a certain sense, for the dependence $\xi(N)$ we have a certain similarity with the results obtained in [16] where a sub-polynomial growth of ξ with N has been found for randomized university networks and preferential attachment models. However, for the RZ3F model the

spectrum has no large gap and is more similar to the real directed networks.

The investigations of RMZ3F model at larger sizes (e.g. with the help of the Arnoldi method [11, 21]) can provide a more firm conclusion about the delocalization properties of eigenstates in this model.

3.4 Properties of AD2 and AD3 models

The spectra of AD2 and AD3 models are shown in Fig. 6a,b with color plot of IPR values. We see that there are rather large values of ξ indicating existence of delocalized eigenstates. Indeed, a more detailed analysis presented in Fig. 7 shows that for the states of the spectral range $\mathrm{Re}\lambda > 0.25$ IPRs are growing with N clearly demonstrating delocalization. Indeed, for maximal ξ from this range (excluding PageRank) we find $\nu = 0.75$ at $d = 2$ and $\nu = 0.95$ at $d = 3$. At the same time in a vicinity of the spectrum edge $\mathrm{Re}\lambda < 0.25$ we have $\nu = 0.18; 0.05$ for $d = 2; 3$ clearly showing that in this part of the spectrum the eigenstates are well localized. Indeed, for these localized states we have an exponential decay $\ln|\psi| \propto -K^{1/d}$ with the eigenstate rank index K (see Fig. 9a,b). Such a decay also appears for the localized states of the Anderson model in dimension d.

But for the majority of eigenstates we have significant growth of ξ with N showing that these states are delocalized. Of course, the case of $d = 2$ should be studied in more detail since for the standard Anderson model at $d = 2$ (3) all eigenstates are exponentially localized [3]. However, we have here non-Hermitian matrix and for our knowledge there are no rigorous results about localization in such matrices in $d = 2$.

Even if in AD2, AD3 models we find delocalization, the PageRank in these models is practically flat due to absence of central node (see Fig. 9a,b). Another weak point of AD2, AD3 models is a relatively narrow distribution of eigenvalues with $|\mathrm{Im}\lambda| < 0.1$ and due to that we continue our analysis with the next model.

3.5 Results for AD2S and AD3S models

The spectra of AD2S, AD3S models are shown in Fig. 6c,d. We see that the additional terms in upper triangle of matrix S produce a broadening of $\mathrm{Im}\lambda$ which however still remains relatively narrow ($|\mathrm{Im}\lambda| < 0.2$). The IPR values are growing with N except of the eigenstates at the spectral edge $\mathrm{Re}\lambda \approx 0.6$ (see Fig. 8). For these localized states the exponent ν is practically zero while for the maximal IPR (except PageRank) we find rather large values of $\nu = 0.57$ at $d = 2$, $\nu = 0.73$ at $d = 3$. Thus, in these models we clearly have the Anderson type transition from localized to delocalized eigenstates.

In analogy with the 3d Anderson model [3], we make a conjecture that in models AD2, AD3, AD2S, AD3S there is a certain mobility edge curve in the complex plane λ which separates localized from delocalized states. In a qualitative manner such a curve is visible in Fig. 6 as a border between blue color of localized states with small ξ and white color of states with large ξ. But definitely more detailed studies are required for a more exact determination of such a mobility edge curve.

Examples of PageRank probability decay are shown in Fig. 9. The new element, appearing in AD2S, AD3S models (comparing to AD2, AD3 cases), is a dependence of the PageRank exponent β on the parameter δ as shown in Fig. 10. These data demonstrate that β increases from $\beta \approx 0.2$ at $\delta = 0.1$ up to $\beta \approx 0.9$ at $\delta = 3$. Thus AD2S,AD3S models have delocalized eigenstates and the PageRank exponent of real directed networks. The only weak point is a narrow distribution of spectrum in $\mathrm{Im}\lambda$. To improve this feature we study in next Section the models AD2Z, AD2ZS.

3.6 Results for AD2Z and AD3ZS models

The spectra of AS2Z, AD2ZS models are shown in Fig. 11. We see that the star structure appears due to introduction of blocks 4×4. The dependence of IPR ξ on $\mathrm{Re}\lambda$ clearly shows the existence of two groups of states with small $\xi < 100$, presumably for localized phase, and large $\xi > 100$, presumably for delocalized phase.

The distribution of ξ on $\lambda-$plane is shown in Fig. 12a,b. Again we see signs of the mobility edge curve separating localized (blue) and delocalized (white) eigenstates.

The dependence of ξ on N is shown in Fig. 12c. There are well localized states with ξ practically independent of N ($\xi < 20$) and delocalized states for which ξ is growing with N with a relatively large growth exponent $\nu = 0.67$ at $\delta = 0$ and $\nu = 0.53$ at $\delta = 0.25$. This gives a strong argument for existence of the Anderson transition with a mobility edge in a complex $\lambda-$plane in these models.

The decay of PageRank probability is shown in Fig. 12d: at $\delta = 0$ we have a flat $P(K)$ distribution with the exponent $\beta = 0.16$, while at $\delta = 0.25$ we find $\beta = 0.51$ being close to the values found in real directed networks (e.g. for the Twitter network $\beta \approx 0.54$ [11]).

Thus we can say that the model AD2ZS is the one being most close to real directed networks with the number of interesting features: algebraic decay of PageRank

probability with the exponent $\beta \approx 0.5$, absence of spectral gap at $\alpha = 1$, a broad star like distribution of eigenvalues in the complex $\lambda-$plane, existence of localized and delocalized eigenstates of the Google matrix with strong indications on the Anderson transition and the mobility curve in $\lambda-$plane.

Of course, the presented results are obtained for matrices of finite size while the Anderson delocalization assumes that states are delocalized over an infinite system size. Due to that the finite matrix simulations give only numerical indications. However, the clear increase of ξ values with matrix size N, shown in Fig. 7, 8, 12, gives us a convincing evidence on existing delocalized states in the limit $N \to \infty$. The same Figs. demonstrate existence of localized states with ξ being practically independent of N.

We expect that a similar model AD3ZS constructed in dimension $d = 3$ from the AD3S model will have even stronger delocalization properties.

4 Summary

Let us now summarize results obtained above in various models. We see that with the Anderson type models without shortcuts (AD2, AD3, AD2Z) we obtain delocalized eigenstates for a certain domain of eigenvalues λ and localized eigenstates at the border of spectrum or at some regions of spectrum. The most interesting model is AD2Z with a broad domain of complex λ (Fig. 12) while for the models AD2, AD3 the imaginary part of λ is relatively small (Fig. 6). For the AD2Z model we have a clear separation between well localized states and those with large ξ corresponding to delocalized domain. A certain curve (a contour) in the complex plain of λ separates these two domains playing the role of mobility edge curve in the complex plain. However, a weak point of model AD2Z is a relatively small value of PageRank exponent $\beta \approx 0.16$. It is possible that more realistic values of $\beta \sim 0.5$ can be reached considering higher dimensions $d = 3, 4, ...$ with the transition blocks used in AD2Z. However, a verification of this expectation requires more advanced numerical simulations in future, probably with the help of the Arnoldi method described in [11].

Another promising case is the AD2ZS model with shortcuts. A small fraction of shortcuts added allows to increase the PageRank exponent and obtain $\beta \approx 0.5$ as in real networks. Even if shortcuts produce transitions to faraway nodes their effect has two tendencies. The first tendency is the delocalization effect since probability can be transferred to nodes located very far. But this does not necessary gives a delocalization since the wave function still can remain located on a small number of nodes with a relatively small ξ values. Such a situation has been seen in the Anderson small world model where a small fraction (density) of shortcuts did not affect localized states remaining with small ξ values and Poisson statistics [18, 19]. The second tendency of shortcuts is related to the fact that they introduce a disorder that may enhance localization features, as it clearly happens in Fig. 5a.

We think that it can be very promising to study the extentions of models AD2Z and AD2ZS to higher dimensions d and larger matrix sizes.

5 Discussion

In this work we described various random matrix models of the Google matrix of directed networks. Our results show that for certain models (like AD2ZS) we have an algebraic decay of PageRank probability with the exponent $\beta \sim 0.5$, absence of spectral gap at $\alpha = 1$, existence of the Anderson transition and mobility edge in the complex $\lambda-$ plane. We think that the further analysis of the models described here will allow to establish more close links between the Anderson delocalization phenomenon in disordered solids and delocalization of eigenstates of the Google matrix of directed networks.

Acknowledgments. We thank L.Ermann and K.M.Frahm for useful discussions. This research is supported in part by the EC FET Open project "New tools and algorithms for directed network analysis" (NADINE *No* 288956). The research of OVZ was partially supported by the Ministry of Education and Science of Russian Federation.

Key words. Markov chains, Anderson localization, Google matrix, PageRank.

References

[1] P. W. Anderson, Phys. Rev. **109**, 1492 (1958).
[2] E. Akkermans and G. Montambaux, Mesoscopic Physics of Electrons and Photons (Cambridge University Press, Cambridge UK, 2007).
[3] F. Evers and A. D. Mirlin, Rev. Mod. Phys. **80**, 1355 (2008).
[4] A. Goetschy and S. E. Skipetrov, arXiv:1303.2880[math-ph] (2013).
[5] S. E. Skipetrov and I. M. Sokolov, Phys. Rev. Lett. **112**, 023905 (2014).
[6] S. Brin and L. Page, Comp. Networks ISDN Sys. **30**, 107 (1998).
[7] A. M. Langville and C. D. Meyer, Google's PageRank and Beyond: The Science of Search Engine Rankings (Princeton University Press, Princeton USA, 2006).

[8] L. Ermann and D. L. Shepelyansky, Acta Phys. Polonica A **120**(6A), A158 (2011).
[9] M. Brin and G. Stuck, Introduction to dynamical systems (Cambridge University Press, Cambridge UK, 2002).
[10] S. Dorogovtsev, Lectures on complex networks (Oxford University Press, Oxford UK, 2010).
[11] L. Ermann, K. Frahm, and D. L. Shepelyansky, arXiv:1409.0428[physics.soc-ph] (2014).
[12] O. V. Zhirov and D. L. Shepelyansky, Phys. Rev. E **81**, 036213 (2010).
[13] L. Ermann and D. L. Shepelyansky, Eur. Phys. J. B **75**, 299 (2010).
[14] L. Ermann and D. L. Shepelyansky, Eur. Phys. J. B **76**, 57 (2010).
[15] K. M. Frahm, Y.-H. Eom, and D. L. Shepelyansky, Phys. Rev. E **89**, 052814 (2014).
[16] O. Giraud, B. Georgeot, and D. L. Shepelyansky, Phys. Rev. E **80**, 026107 (2009).
[17] K. Zyczkowski, M. Kus, W. Slomczynski, and H.- J. Sommers, J. Phys. A: Math. Gen **36**, 3425 (2003).
[18] A. D. Chepelianskii and D. L. Shepelyansky, http://www.quantware.ups-tlse.fr/talks-posters/chepelianskii2001.pdf (2001).
[19] O. Giraud, B. Georgeot, and D. L. Shepelyansky, Phys. Rev. E **72**, 036203 (2005).
[20] K. M. Frahm, A. D. Chepelianskii, and D. L. Shepelyansky, J. Phys. A: Math. Theor. **45**, 405101 (2012).
[21] K. M. Frahm, B. Georgeot, and D. L. Shepelyansky, J. Phys. A: Math. Theor. **44**, 465101 (2011).
[22] L. Ermann, K. M. Frahm, and D. L. Shepelyansky, Eur. Phys. J. B **86**, 193 (2013).
[23] Ya. G. Sinai, Theor. Prob. Appl. **27**(2), 247 (1982).
[24] P. Le Doussal, J. Stat. Mech: Theor. Exp. **2009**(7), P07032 (2009).