Google matrix, dynamical attractors, and Ulam networks

D. L. Shepelyansky[1,2] and O. V. Zhirov[3,2]
[1]*Laboratoire de Physique Théorique (IRSAMC), Université de Toulouse–UPS, F-31062 Toulouse, France*
[2]*LPT (IRSAMC), CNRS, F-31062 Toulouse, France*
[3]*Budker Institute of Nuclear Physics, 630090 Novosibirsk, Russia*
(Received 26 May 2009; revised manuscript received 20 August 2009; published 16 March 2010)

We study the properties of the Google matrix generated by a coarse-grained Perron-Frobenius operator of the Chirikov typical map with dissipation. The finite-size matrix approximant of this operator is constructed by the Ulam method. This method applied to the simple dynamical model generates directed Ulam networks with approximate scale-free scaling and characteristics being in certain features similar to those of the world wide web with approximate scale-free degree distributions as well as two characteristics similar to the web: a power-law decay in PageRank that mirrors the decay of PageRank on the world wide web and a sensitivity to the value α in PageRank. The simple dynamical attractors play here the role of popular websites with a strong concentration of PageRank. A variation in the Google parameter α or other parameters of the dynamical map can drive the PageRank of the Google matrix to a delocalized phase with a strange attractor where the Google search becomes inefficient.

DOI: 10.1103/PhysRevE.81.036213 PACS number(s): 05.45.Ac, 89.20.Hh

I. INTRODUCTION

The world wide web (WWW) continues its striking expansion going beyond 10^{11} web pages [1]. Information retrieval from such an enormous database becomes the main challenge for WWW users. An efficient solution, known as the PageRank algorithm (PRA) proposed by Brin and Page in 1998 [2], forms the basis of the Google search engine used by the majority of internet users in everyday life. The PRA is based on the construction of the Google matrix which can be written as (see, e.g., [3] for details):

$$\mathbf{G} = \alpha \mathbf{S} + (1 - \alpha)\mathbf{E}/N. \tag{1}$$

Here the matrix $\mathbf{S}$ is constructed from the adjacency matrix $\mathbf{A}$ of directed network links between N nodes so that $S_{ij} = A_{ij}/\Sigma_k A_{kj}$ and the elements of columns with only zero elements are replaced by $1/N$. The second term on the right-hand side (rhs) of Eq. (1) describes a finite probability $1 - \alpha$ for a WWW surfer to jump at random to any node so that $E_{ij}=1$. This term stabilizes the convergence of the PRA by introducing a gap between the maximal eigenvalue $\lambda=1$ and the other eigenvalues λ_i. Usually the Google search uses the value $\alpha=0.85$ [3]. By construction $\Sigma_i G_{ij}=1$ so that the asymmetric matrix $\mathbf{G}$ has a left eigenvector that is a homogeneous constant for $\lambda=1$. The right eigenvector at $\lambda=1$ is the PageRank vector with positive elements p_j and $\Sigma_j p_j=1$ (the components of this vector give the values p_j). All WWW nodes can be ordered by decreasing p_j so that the PageRank plays a role in the ordering of websites and information retrieval. The classification of nodes in terms of the decreasing order of p_j values is used to classify importance of network nodes, as it is described in more detail in [3] (at least this is an important element of the WWW ranking used by Google but additional criteria are also used). The information retrieval and ordering is based on this classification and we also use it in the following.

It is interesting and important to note that by construction the operator $\mathbf{G}$ belongs to the class of Perron-Frobenius operators [3]. Such type operators naturally appear in ergodic theory [4] and in the description of dynamical systems with Hamiltonian or dissipative dynamics [5,6].

The numerical studies of properties of $\mathbf{G}$ are usually done only for the PageRank vector which can be found efficiently by the PRA due to a relatively small average number of links in the WWW (we have not found numerical data on the spectrum of $\mathbf{G}$ in the available literature, see however [3,7–11]). It is established that for large WWW subsets p_j is satisfactory described by a scale-free algebraic decay with $p_j \sim 1/j^{\beta}$, where j is the PageRank ordering index and $\beta \approx 0.9$ [3,7]. Studies of PageRank properties are now very common within the computer science community and they are presented in a number of interesting publications (see, e.g., [8–10] and an overview of the field in [11]). Also a number of rigorous mathematical results have been obtained in this field (see, e.g., [11,12]).

While the properties of the PageRank vector are of primary importance, it is also interesting to analyze the properties of the Google matrix $\mathbf{G}$ as a whole large matrix. Such an analysis can help to establish links between the Google matrix and other fields of physics where large matrices play an important role. Among such fields we can mention the random matrix theory [13] which finds applications in a description of spectra in complex many-body quantum systems and Anderson localization, which is an important physical phenomenon for electron transport in disordered systems (see, e.g., [14]). A transition from localized to delocalized eigenstates also can take place in networks of small world type (see [15,16]). However, in the physical systems considered in [13–16] all matrices are Hermitian with real eigenvalues, while the Perron-Frobenius matrices have generally complex eigenvalues.

A recent attempt to analyze numerically the properties of right eigenvectors ψ_i ($\mathbf{G}\psi_i=\lambda_i\psi_i$) and complex eigenvalues λ_i was done in [17]. The Google matrix was constructed from a directed network generated by the Albert-Barabasi model and the WWW University networks with randomization of links. The Google matrix was considered mainly for

the value $\alpha=0.85$. It was shown that under certain conditions a delocalized phase emerges for the PageRank and other states having complex λ. In spite of a number of interesting results found in [17] a weak feature of the models used there is a significant gap between $\lambda=1$ of the PageRank vector of matrix **S** (or matrix **G** at $\alpha=1$) and $|\lambda_i|\leq 0.4$ for the other vectors. We note that according to [17] the University networks have $|\lambda_i|$ close to 1 but after randomization of links a large gap emerges in the spectrum of λ. This gap in $|\lambda|$ was rather large and hence the PageRank vector itself was not very sensitive to α. In contrast, for the real WWW it is know that p_j is rather sensitive to α due to the existence of $|\lambda_i|$ close to 1 [3,17]. Thus the results obtained in [17] show that even if the Google matrix is constructed on the basis of typical models of scale-free networks, it is quite possible that its spectrum may have a large gap for $0.85\leq\alpha\leq 1$ (thus being rather far from spectral properties of the Google matrices of the WWW). Therefore, it is rather desirable to have other simple models which generate a directed network with Google matrix properties being close to those of WWW.

With an aim to having more realistic models, we develop in this work another approach and construct the Google matrix from the Perron-Frobenius operator generated by a certain dynamical system. The probability flow in dynamical system models has rich and nontrivial features of general importance such as simple and strange attractors with localized and delocalized dynamics governed by simple dynamical rules. Such objects are generic for nonlinear dissipative dynamics and hence can have relevance for actual WWW structure. Thus these objects can find some reflections in the PageRank properties. The dynamical system is described by the Chirikov typical map [18] with dissipation, the properties of this simple model have been analyzed in detail in a recent work [19]. We find that the Google matrix generated by this dynamical model has many λ_i close to 1 and the PageRank becomes sensitive to α (see Fig. 1). This model also captures other specific properties of the WWW Google matrices such as an approximate power-law degree distribution and PageRank decay close to the web (see Figs. 4 and 6).

To construct a network of nodes from a continuous two-dimensional phase space, we divide the space of dynamical variables (x,y) on $N=N_x\times N_y$ cells (we use $N_x=N_y$). Then N_c trajectories are propagated from a cell j over the whole period of the dynamical map and the elements S_{ij} are taken to be equal to a relative number N_i of trajectories arrived at a cell i ($S_{ij}=N_i/N_c$ and $\Sigma_i S_{ij}=1$). Thus **S** gives a coarse-grained approximation of the Perron-Frobenius operator for the dynamical map. The Google matrix **G** of size N is constructed from **S** according to Eq. (1). We use sufficiently large values of N_c so that the properties of **G** become insensitive to N_c.

Such a discrete approximation of the Perron-Frobenius operator is known in dynamical systems as the Ulam method [20]. Indeed, Ulam conjectured that such a matrix approximant correctly describes the Perron-Frobenius operator of continuous phase space. For hyperbolic maps the Ulam conjecture was proven in [21]. Various types of more generic one-dimensional maps have been studied in [22–24]. Further mathematical results have been reported in [25–28] with extensions and proof of convergence for hyperbolic maps in higher dimensions. However, the studies of more generic two-dimensional maps remain rather restricted (see, e.g., [29]) and nonsystematic. In principle the construction of directed graphs on the basis of dynamical systems is a known mathematical approach (see, e.g., [6]) but the spectral properties of the Google matrix built on such graphs were not studied until now.

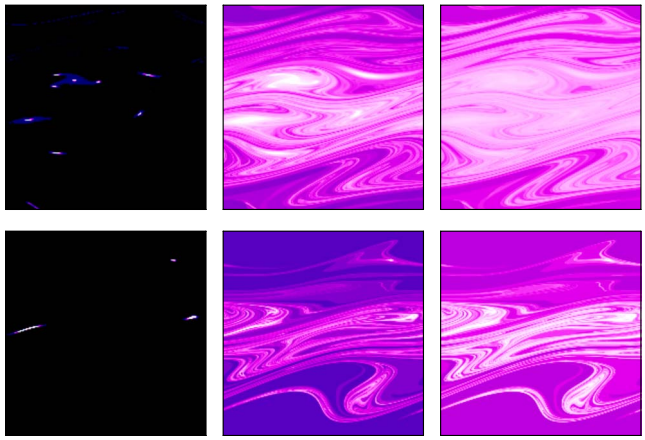

FIG. 1. (Color online) PageRank p_j for the Google matrix generated by the Chirikov typical map (2) at $T=10$, $k=0.22$, $\eta=0.99$ (set $T10$, top row), and $T=20$, $k=0.3$, $\eta=0.97$ (set $T20$, bottom row) with $\alpha=1$ (left column), 0.95 (middle column), 0.85 (right column); the value of index j is a function of the cell index in the map phase space $(i_x;i_y)$ (see text), here p_j is plotted in phase space of the map using the above relation between j and (i_x,i_y). The phase space region $0\leq x<2\pi;-\pi\leq y<\pi$ is divided on $N=3.6\times 10^5$ cells; p_j is zero for black and maximal for white.

Of course, one can construct other type of models of the Google matrix (see, e.g., copying models known in computer science [30]). However, the advantage of the Ulam networks is based on the deep mathematical and physical studies performed on the dynamical systems over the last few decades (see, e.g., [4,31,32]).

In this paper we show that the Ulam method applied to two-dimensional dissipative dynamical maps generates a type of directed networks which we call the Ulam networks. We present here numerical and analytical studies of certain properties of the Google matrix of such networks.

The paper is organized as follows: in Sec. II we give the description of the Chirikov typical map and the way the Ulam network is constructed on the basis of this map with the corresponding Google matrix; the properties of this map and network are also described here; in Sec. III the properties of the eigenvalues and eigenstates of the Google matrix are analyzed in detail, including the delocalization transition for the PageRank, the fractal Weyl law, and the global contraction properties; a summary of the results is presented in Sec. IV.

II. ULAM NETWORKS OF DYNAMICAL MAPS

A. Chirikov typical map

To construct an Ulam network and a generated by it Google matrix we use a dynamical two-dimensional dissipa-

tive map. The dynamical system is described by the Chirikov typical map introduced in 1969 for a description of continuous chaotic flows [18]:

$$y_{t+1} = \eta y_t + k \sin(x_t + \theta_t), \quad x_{t+1} = x_t + y_{t+1}. \tag{2}$$

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. 1 ($0 \le x < 2\pi$, $-\pi \le y < \pi$) with the 2π-periodic boundary conditions. The parameter $0 < \eta \le 1$ gives the global dissipation. The properties of the symplectic map at $\eta = 1$ have been studied recently in detail [19]. The dynamics is globally chaotic for $k > k_c \approx 2.5/T^{3/2}$ and the Kolmogorov-Sinai entropy is $h \approx 0.29k^{2/3}$ (more details about chaotic dynamics and the Kolmogorov-Sinai entropy can be found in [4,5,31,32]).

In this study we use two random sets of phases θ_t with $T=10$ and $T=20$. Their values are given in the Appendix. We also fixed the dissipation parameter $\eta=0.99$ for $T=10$ and $\eta=0.97$ for $T=20$. We call these two sets of parameters as $T10$ and $T20$ sets, respectively. The majority of data are obtained at $k=0.22$ for the set $T10$ and at $k=0.3$ for the set $T20$ (see Fig. 1). These are two main working points for this work.

For the set $T10$ ($k=0.22$, $\eta=0.99$) we have the theoretical value of the Kolmogorov-Sinai entropy $h=0.29k^{2/3}=0.105$ for the symplectic map at $\eta=1$ [19]. The actual value at $\eta=1$ is determined numerically by the computation of the Lyapunov exponent and has a value $h=0.0851$. For $\eta=0.99$ we also have the global dissipation rate $\gamma_c = -T \ln \eta = 0.1005$ after the map period (which is equal to T iterations). The global contraction factor is $\Gamma_c = \eta^T = \exp(-\gamma_c) = 0.9043$. For a weak dissipation the fractal dimension d of the limiting set can be approximately estimated in a usual way (see, e.g., [32]) as $d = 2 - \gamma_c/(Th) = 1.882$.

In a similar way for the set $T20$ ($k=0.3$, $\eta=0.97$) we have the theoretical value $h=0.29k^{2/3}=0.1299$, while the actual numerical value is $h=0.1081$. Also here $\gamma_c = -T \ln \eta = 0.609$, $\Gamma_c = 0.5437$, and the estimated fractal dimension of the limiting set is $d = 2 - \gamma_c/(Th) = 1.718$.

The bifurcation diagrams for the sets $T10$ and $T20$ are shown in Figs. 2 and 3, respectively. On large time scales we clearly see parameter k regions with simple and chaotic attractors. For a shorter time scales a distinction between two regimes becomes less pronounced. This means that during a long time a trajectory moves between few simple attractors (which are represented by short periodic orbits and are clearly seen in Fig. 1 in the left column) before a final convergence is reached.

B. Network construction and distribution of links

The Ulam network for the Chirikov typical map [Eq. (2)] is constructed in the following way. The whole phase space region $2\pi \times 2\pi$ is divided into $N = N_x \times N_y$ cells ($N_x = N_y$) and N_c trajectories are propagated from each given cell j during

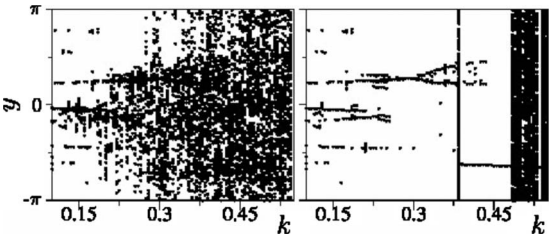

FIG. 2. Bifurcation diagram showing values of y vs map parameter k for the set $T10$ of the Chirikov typical map [Eq. (2)]. The values of y, obtained from ten trajectories with initial random positions in the phase space region, are shown for integer moments of time $100 < t/T \le 110$ (left panel) and $10^4 < t/T \le 10^4 + 100$ (right panel).

T map iterations which form the period of the map. After that the elements of matrix S_{ij} are computed as $S_{ij} = N_i/N_c(j)$, where N_i is a number of trajectories arrived from a cell j to cell i. In this way we have by a definition $\Sigma_i S_{ij} = 1$. Such $\mathbf{S}$ gives a coarse-grained approximation of the Perron-Frobenius operator for the map [Eq. (2)]. The Google matrix $\mathbf{G}$ of size N is constructed from $\mathbf{S}$ according to Eq. (1). To construct S_{ij} we usually use $N_c = 10^4$ but the properties of $\mathbf{S}$ are not affected by a variation in N_c in the interval $10^3 \le N_c \le 10^5$. Since the cell size is very small it is unimportant in what way N_c trajectories are distributed inside the cell. The index j of the PageRank vector p_j is a function of cell indexes (i_x, i_y) which is used for the phase space plot of p_j in Fig. 1. Up to statistical fluctuations, the values of S_{ij} remain the same for homogeneous or random distribution of N_c trajectories inside a cell. We note that in our system we have only one nondegenerate eigenvalue $\lambda = 1$ of $\mathbf{G}$ at $\alpha = 1$ with the corresponding eigenvector. The standard numerical methods find this value without any problem for moderate sizes N. However, at very large N the quasidegeneracy with other eigenvalues, which approach to unity exponentially with N [see Fig. 7(a) below in the text], can go beyond the numerical accuracy of diagonalization. We note that in real WWW there are certain exact degeneracies of $\lambda = 1$ eigenvalue which we believe are linked with a small number of links and thus a small number of different matrix elements in $\mathbf{S}$ matrix (see, e.g., [11,12,33]). In the Ulam networks the matrix elements fluctuate with the number of trajectories N_c and there are practically no matrix elements with exactly the same values. In any case for the matrix sizes shown in Fig. 7(a) we have quasidegeneracy near $\lambda = 1$ at $\alpha = 1$, while for the same sizes of the university networks, considered in [17,33], we obtain exact degeneracy of $\lambda = 1$ eigenvalue (up to a computer precision). Thus the structure of degeneracy of $\lambda = 1$ is different in both types of networks.

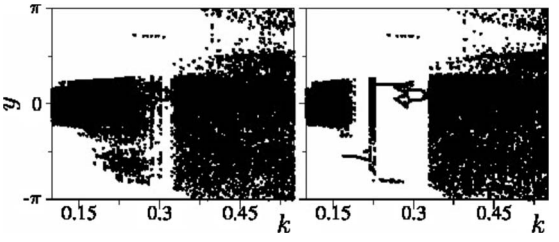

FIG. 3. Same as in Fig. 2 for the set $T20$.

Up to $N=22\,500$ we used exact diagonalization of **G** to determine all eigenvalues λ_j and right eigenvectors ψ_i, for larger N up to $N=1.44\times10^6$ we used the PRA to determine the PageRank vector. The majority of data are presented for two typical sets $T10$ and $T20$ of parameters of the map [Eq. (2)] and the PageRanks for various values of α are shown in Fig. 1. For these sets the dynamics has a few fixed point attractors but it takes a long time $t\sim10^3$ to reach them. During this time a trajectory visits various regions of phase space.

It is important to note that the discreteness of phase space, linked to a finite cell size, produces an important physical effect which is absent in the original continuous map [Eq. (2)]: effectively it introduces an additional noise which amplitude σ is approximately $\sigma\sim2\pi/\sqrt{N}$. This becomes especially clear for the symplectic case at $\eta=1$ and at small values of k at $T=1$ (all θ_t are the same). In this case the map is reduced to the Chirikov standard map [31] and the continuous map dynamics is bounded by the invariant Kolmogorov-Arnold-Moser (KAM) curves. However, the discreteness of phase space allows jump from one cell to another and thus to go from one invariant curve to another one. This leads to a diffusion in y and appearance of a homogeneous ergodic state at $\lambda=1$. A direct analysis also shows that at any finite cell size the operator **S** has a homogeneous ergodic state with $\lambda=1$, we also checked this via numerical diagonalization of matrix sizes $N\approx20\,000$. This example shows that the Ulam conjecture is not valid for quasi-integrable symplectic maps in the KAM regime.

The physical origin of the difference between the continuous map and the finite-size cell approximation is due to introduction of an effective noise term σ_t in rhs of Eq. (2) induced by a finite cell size. Due to this noise the trajectories diffuse over all region $-\pi<y<\pi$ after a diffusive time scale $t_D\sim\pi^2/\sigma^2$ even if the continuous map is in the KAM regime with bounded dynamics in y. Hence, here $\sigma\sim2\pi/\sqrt{N}$ is an effective amplitude of noise introduced by cell discreteness.

Even if this σ-noise leads to a drastic change in dynamics for quasi-integrable regime its effects are not very important in the case of chaotic dynamics where noise gives only a small additional variation as compared to strong dynamical variations induced by dynamical chaos. With such a physical understanding of discreteness effects we continue to investigate the properties of the Ulam networks. However, we stress that the σ-noise is local in the phase space and hence it is qualitatively different from the Google term α which generates stochastic jumps over all sites.

In Figs. 4 and 5 we show the distributions of ingoing $P_{in}(\kappa)$ and outgoing $P_{out}(\kappa)$ links κ in the Ulam network presented by **S** matrix generated by the map [Eq. (2)] as described above. These distributions are satisfactory described by a scale-free algebraic decay $P\sim1/\kappa^{\mu}$ with $\mu\approx1.86$, 1.11 for ingoing and 1.91, 1.46 outgoing links at $T10$ and $T20$, respectively, and a typical number of links per node $\kappa\sim10$ (see Figs. 4 and 5). Such values are compatible with the WWW data of scale-free type where $\mu\approx2.1$, 2.7 for ingoing, outgoing links [3,7]. However, we may also note an appearance of certain deviations at large values of κ. Indeed, for a dynamical system a large number of links appears due to exponential stretching of one cell after T map iterations

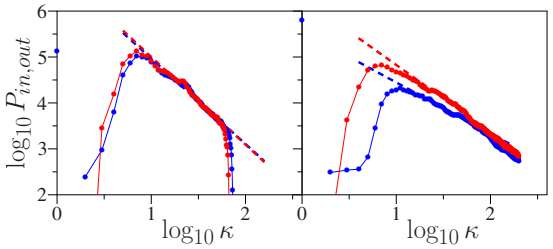

FIG. 4. (Color online) Differential distribution of number of nodes with *ingoing* $P_{in}(\kappa)$ (blue/black) and *outgoing* $P_{out}(\kappa)$ (red/gray) links κ for sets $T10$ (left) and $T20$ (right) (top and bottom curves at $\log_{10}\kappa=0.6$ correspond to red/gray and blue/black curves). The straight dashed lines give the algebraic fit $P(\kappa)\sim\kappa^{-\mu}$ with the exponent $\mu=1.86$, 1.11 ($T10$,$T20$) for *ingoing* and $\mu=1.91$, 1.46 ($T10$,$T20$) *outgoing* links. Here $N=1.44\times10^6$ and $P(\kappa)$ gives a number of nodes at a given integer number of links κ for this matrix size. Isolated blue/black point at $\kappa=0$ shows that in the whole matrix there is a significant number of nodes with zero *ingoing* links. We note that the WWW is characterized by the exponents $\mu\approx2.1$, 2.7 for ingoing and outgoing links respectively (see, e.g., [3,7]).

that gives a typical number of links $k\sim\exp(hT)$. It is possible that during the dynamical evolution much larger values of stretching can appear. Indeed, the comparison of two cases at $k=0.22$ and $k=0.6$ for the set $T10$ in Fig. 5 shows that for larger k the scale-free distribution continues to much larger values of $\kappa>200$ while for smaller k the scale-free type decay stops around $\kappa\approx50$. For the set $T20$ the stretching is stronger and the scale-free decay continues up to larger values of κ.

It is clear that for the Ulam networks discussed here one has a rapid exponential decay of links distribution at asymptotically large link number κ. However, due to an exponential growth of typical $\kappa\sim\exp(hT)$ a scale-free-type decay can be realized up to very large κ by increasing T. In these studies we stay at the chosen working points where a scale-free decay remains dominant for matrix sizes of the order of $N\sim10^5$–10^6. We note that in our model the algebraic decay of links distribution starts from finite values of $j\approx7$, while in

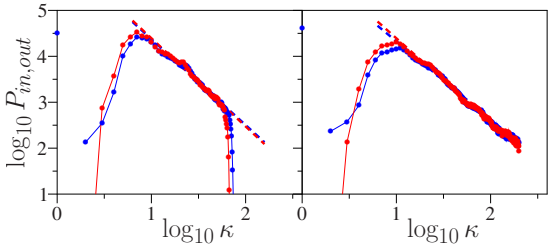

FIG. 5. (Color online) Same as in Fig. 3 for the set $T10$ at $k=0.22$ (left) (same as Fig. 4 left) and $k=0.6$ (right) and $N=3.6\times10^5$. The fit gives the exponent $\mu=1.87$, 1.92 for *ingoing* (blue/black), *outgoing* (red/gray) links at $k=0.22$ (left) and $\mu=1.70$, 1.83 for *ingoing* (blue), *outgoing* (red) links at $k=0.6$ (right) (top and bottom curves at $\log_{10}\kappa=0.8$ correspond to red/gray and blue/black curves). Isolated blue/black point at $\kappa=0$ shows that in the whole matrix there is a significant number of nodes with zero *ingoing* links.

the WWW networks it starts from $j \approx 1$. However, the main interest is to the properties at large j where we have a large range of algebraic decay of links distribution.

Finally we note that the models of the Google matrix generated by the Ulam networks are most interesting for dissipative maps. Indeed, by construction the left eigenvector of the Google matrix $\psi_i^+ \mathbf{G} = \psi_i^+$ at $\lambda = 1$ is a homogeneous vector $\psi_i^+ = \text{const}$ As a result for symplectic maps the right vector of PageRank p_j is also homogeneous. Only dissipation term generates an inhomogeneous decay of p_j.

III. PROPERTIES OF EIGENVALUES AND EIGENSTATES

A. Delocalization transition for PageRank with α

The variation in PageRank p_j with α is shown in Fig. 1 for two sets $T10$ and $T20$. The distribution p_j is plotted for each cell of the phase space (x,y), the numbering of cells is done by the integer grid $n_x \times n_y$ which has a certain correspondence with the index j which numerates the values of p_j in the decreasing order with j. At $\alpha = 1$ the distribution p_j is concentrated only on a few local spots corresponding to fixed point attractors. Physically this happens due to presence of σ noise, induced by cell discretization, which leads to transitions between various fixed points. With the decrease in α the PageRank starts to spread over a strange attractor set. The properties of strange attractors in dynamical dissipative systems are described in [32]. In the map [Eq. (2)] the strange attractor appears at larger values of k (namely, $k > 0.5$ for $T10$ and $k > 0.34$ for $T20$, see Figs. 2 and 3) but a presence of effective noise induced by σ and $1-\alpha$ terms leads to an earlier emergence of strange attractor. Below a certain value $\alpha < \alpha_c$ the PageRank becomes completely delocalized over the strange attractor as it is clearly seen in Fig. 1 for the set $T10$.

The dependence of p_j on j is shown in more detail in Fig. 6. For $\alpha = 1$ PageRank shows a rapid drop with j that can be fitted by an exponential Boltzmann type distribution $p_j \sim \exp(-b\gamma_c j / D_\sigma)$, where b is a numerical constant ($b \approx 1.4$; 2.1 for $T10$; $T20$), $\gamma_c = -T \ln \eta$ is the global dissipation rate, and $D_\sigma = \sigma^2 N \approx (2\pi)^2$ is σ noise diffusion [dashed lines in Figs. 6(a) and 6(d)]. Such an exponential decay results from the Fokker-Planck description of the map [Eq. (2)] in the presence of σ noise term which gives diffusive transitions on nearby cells. For $\alpha < 1$, transitions of a random surfer, introduced by Google, give a significant modification of PageRank which shows an algebraic decay $p_j \sim 1/j^\beta$ with the exponent β dependent on α [Figs. 6(b)–6(f)]; for the set $T20$ at $\alpha = 0.95$ we obtain $\beta \approx 0.88$ being close to the numerical value found for the WWW [3,7] which has $\beta \approx 0.9$. This fact shows that the Ulam networks can model certain properties of the WWW, but one should keep in mind existing significant differences between these two networks (e.g., in our Ulam network we have exponential decay of links distribution at large number of links). However, β decreases with the decrease in α and for $T10$ set a delocalization takes place for $\alpha = 0.85$ so that p_j spreads homogeneously over the strange attractor [see Fig. 1 top right panel and Fig. 6(c)]. For $T20$ set $p_j \sim \psi_{i=1}(j)$ remains localized at $\alpha = 0.85$. To define the localization in a more quantitative way we use a participation ratio (PAR) defined for the eigenstate $\psi_i(j)$ as $\xi_i = (\Sigma_j |\psi_i(j)|^2)^2 / \Sigma_j |\psi_i(j)|^4$. We define that the PageRank is localized if its ξ remains finite at large N. We use this definition of PAR ξ for all eigenvectors $\psi_i(j)$ omitting the index i. This quantity PAR is broadly used in solid state systems with disorder (see, e.g., [14–17]); ξ gives an effective number of nodes populated by an eigenstate.

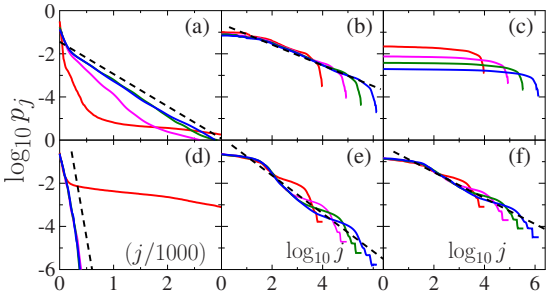

FIG. 6. (Color online) PageRank distribution p_j for $N = 10^4$, 9×10^4, 3.6×10^5, and 1.44×10^6 (larger N have more dark and more long curves in (b), (c), (e), and (f); in (a) this order of N is for curves from bottom to top (curves for $N = 3.6 \times 10^5$ and 1.44×10^6 practically coincide in this panel) and in (d) this order of N is for curves from top to bottom at $j \approx 500$ (curves for $N = 9 \times 10^4$, 3.6×10^5, 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, and blue curves, respectively). The dashed straight lines show fits $p_j \sim 1/j^\beta$ with β: (b) 0.48, (e) 0.88, and (f) 0.60. Dashed lines in panels (a) and (d) show an exponential Boltzmann decay (see text, lines are shifted in j for clarity). Other parameters, including the values of α, and panel order are as in Fig. 1. In panels (a) and (d) the curves at large N become superimposed. Here and below logarithms are decimal.

B. Properties of other eigenvectors

To understand the origin of the delocalization transition in α we analyze in Fig. 7 the properties of all eigenvalues λ_i and eigenvectors ψ_i with their PAR ξ. Due to σ noise activation transitions take place between the attractor fixed points leading to states with λ_i being exponentially close to $\lambda = 1$ [Fig. 7(a)]. The convergence to $|\lambda| = 1$ is exponential in N for certain states and may lead to numerical problems at very large N. However, the standard numerical diagonalization methods remained stable for the values of N used in our studies.

The distribution of λ_i in the complex plane is shown in Figs. 7(c) and 7(d): there are λ_i approaching $\lambda = 1$ mainly along the real axis but a majority of λ_i are distributed inside a circle of finite radius around $\lambda = 0$; this radius decreases with the increase in global dissipation from $\gamma_c = 0.10$ for set $T10$ to $\gamma_c = 0.61$ for $T20$. The PAR values for states inside the circle have typical values $4 \leq \xi \leq 300$ shown by grayness. The dependence of ξ on $\gamma = -2 \ln|\lambda|$ and N shows that the eigenstates inside the circle remain localized at large N [Fig. 7(b)]. We attribute this to the fact that at large N the diffusion due to σ noise in presence of dissipation leads to spreading only over a finite number of cells and thus ξ remains bounded. This $\xi(\gamma, N)$ dependence is different from one ob-

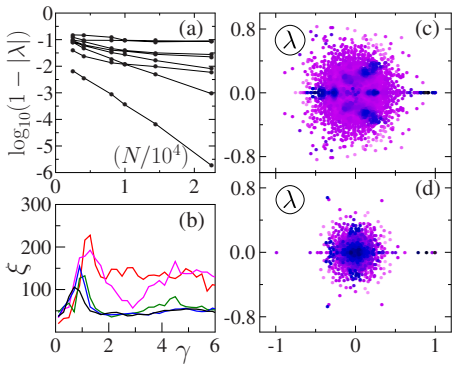

FIG. 7. (Color online) (a) Dependence of gap $1-|\lambda|$ on Google matrix size N for few eigenstates with $|\lambda|$ most close to 1, set $T10$, $\alpha=1$; (b) dependence of PAR ξ on $\gamma=-2\ln|\lambda|$ for $N=2500$, 5625, 8100, 10^4, 14 400 for set $T10$, $\alpha=1$ (curves from top to bottom at $\gamma=4$, in online version this order of curves corresponds to colors red, magenta, green, blue, black); (c) complex plane of eigenvalues λ for set $T10$ with their PAR ξ values shown by grayness (black/blue for minimal $\xi\approx 4$, gray/light magenta for maximal $\xi\approx 300$; here $\alpha=1$, $N=1.44\times 10^4$); (d) same as (c) but for set $T20$.

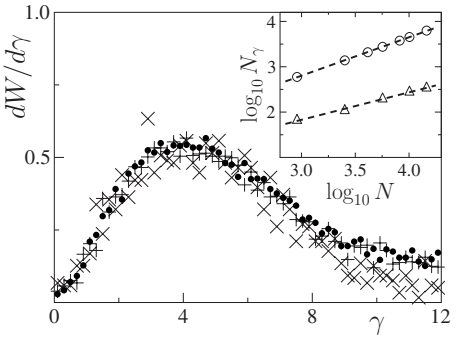

FIG. 8. Probability distribution $dW(\gamma)/d\gamma$ for set $T10$, $\alpha=1$ at $N=2.5\times 10^3(\times)$, $10^4(+)$, 1.44×10^4 (dots); $W(\gamma)$ is normalized by the number of states $N_\gamma=0.55N^{0.85}$ with $\gamma<6$. Inset: dependence of number of states N_γ with $\gamma<\gamma_b$ on N for sets $T10$ (circles, $\gamma_b=6$) and $T20$ (triangles, $\gamma_b=3$); dashed lines show the fit $N_\gamma=AN^\nu$ with $A=0.55$, $\nu=0.85$ and $A=0.97$, $\nu=0.61$ respectively.

tained in [17] for the Albert-Barabasi model, the comparison with data from WWW University networks is less conclusive due to strong fluctuations from one network to another (see Fig. 4 in [17]): an average growth of ξ is visible there even if at $N\sim 10^4$ the values of ξ are comparable with those of Fig. 7(b). Globally our data of Fig. 7 show that the diffusive modes at $|\lambda_i|<1$ remain localized on a number of nodes $\xi \ll N$.

We also stress an important property of eigenvalues and eigenvectors with $0<|\lambda_i|<1$. Our numerical data show that for the states with $0<|\lambda_i|<1$ their ξ_i are independent of α (λ_i are simply rescaled by a factor α according to [3]). This happens due to a specific property of $(1-\alpha)\mathbf{E}/N$ term in $\mathbf{G}$, which is constructed from a homogeneous vector with rank equal to unity. Right eigenvectors are orthogonal to the homogeneous left vector and hence $(1-\alpha)$ term affects only the PageRank but not other eigenvectors. This property can be easily obtained from the theorems for eigenvalues dependence on *alpha* presented in [3].

C. Fractal Weyl law for Google matrix

Another interesting characteristic of $\mathbf{G}$ at $\alpha=1$ is the distribution of eigenvalues in the complex plain. In this subsection we study the density distribution $dW(\gamma)/d\gamma$ over γ linked to the absolute value $|\lambda|$. Here $dW(\gamma)$ gives the number of states in the interval $d\gamma$ with a certain global numerical normalization factor (see Fig. 8). The data presented in Fig. 8 show that its form becomes size independent in the limit of large N. At small $\gamma<3$ the density decreases approximately linearly with γ without any large gap. We find that the total number of states N_γ with finite $\gamma<\gamma_b\approx 5$ grows algebraically as $N_\gamma=AN^\nu$ with $\nu<1$ (Fig. 8 inset). We interpret this result on the basis of the fractal Weyl law established recently for nonunitary matrices with fractal eigenstates (see, e.g., [34,35] and references therein). According to this law the exponent ν is $\nu=d-1$ where d is the fractal dimension of the dynamical system. Approximately we have $d-1\approx 1-\gamma_c/(Th)$ [32,35] that gives $\nu=0.88$, 0.72 for the sets $T10$ and $T20$ with the numerical values of γ_c, h given above. These values are in a good agreement with the fit data $\nu=0.85$, 0.61 of Fig. 8 inset. The fact that $\nu<1$ implies that almost all states have $\lambda=0$ in the limit of large N (in this work we do not discuss the properties of these degenerate states with large $\xi\sim N$).

It is interesting to note that the fractal Weyl law is usually discussed for the open quantum chaos systems (see [34,35] and references therein). There the matrix size is inversely proportional to an effective Planck constant $N\propto 1/\hbar$. For the Ulam networks generated by dynamical attractors a cell size in the phase space places the role of effective $\hbar$. This opens interesting parallels between quantum chaotic scattering and discrete matrix representation of the Perron-Frobenius operators of dynamical systems.

D. PageRank delocalization again

The dependence of PAR ξ of the PageRank on α and N is shown in Fig. 9. It permits to determine the critical value α_c below which PageRank becomes delocalized showing ξ growing with N. According to this definition we have ξ independent of large N for $\alpha>\alpha_c$, while for $\alpha<\alpha_c$ the PAR ξ grows with N. The obtained data give $\alpha_c\approx 0.95$, 0.8 for $T10$ and $T20$. Further investigations are needed to understand the dependence of α_c on system parameters. Here we make a conjecture that $1-\alpha_c\approx C\gamma_c\ll 1$ with a numerical constant $C\approx 0.3$. Indeed, for larger dissipation rate $\gamma_c=-T\ln\eta$ a radius of a circle with large density of λ_i in the complex plane λ becomes smaller [see Figs. 7(c) and 7(d)] and thus larger values of $1-\alpha$ are required to have a significant contribution of these excited relaxation modes to the PageRank. Also the data of [35] for systems with absorption rate γ_c show a low density of states at $\gamma<\gamma_c$ so that it is natural to expect that one should have $1-\alpha_c\sim\gamma_c$ to get a significant contribution of delocalized relaxation modes from a strange attractor set to the PageRank. It is quite probable that C depends in addition on system parameters. Indeed, even at fixed γ_c and $\alpha=0.99$ being rather close to 1 it is possible to have a tran-

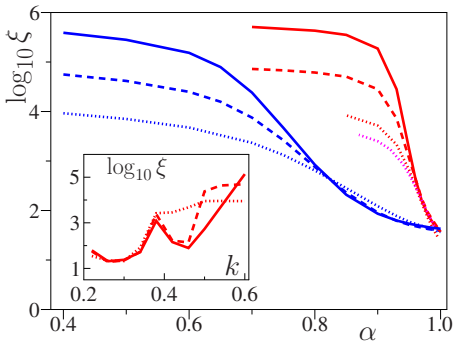

FIG. 9. (Color online) Dependence of PageRank ξ on α for set $T10$ at $N=5625$ (dotted magenta/gray), 1.44×10^4 (dotted red/gray, 9×10^4 (dashed red/gray, 6.4×10^5 (full red/gray), (these four curves follow from bottom to top at $\alpha=0.9$ for small to large values of N respectively); and for set $T20$ at $N=1.44\times10^4$ (dotted blue/black), 9×10^4 (dashed blue/black), 6.4×10^5 (full blue/black), (these three blue/black curves follow from bottom to top at $\alpha=0.6$). Inset shows dependence of ξ on k for set $T10$ at $\alpha=0.99$ with $N=1.44\times10^4$ (dotted red/gray), 9×10^4 (dashed red/gray), 3.6×10^5 (full red/gray) (this corresponds to the order of curves from top to bottom at $k=0.45$).

sition from localized to delocalized PageRank by increasing k in the map [Eq. (2)] (see Fig. 9 inset and Fig. 10). This transition in k takes place approximately at $k\approx0.55$ when fixed point attractors merge into a strange attractor (see the bifurcation diagram in Fig. 2). A peak in ξ around $k\approx0.38$ is related to birth and disappearance of a strange attractor in a narrow interval of k at $k\approx0.38$. At the same time an increase in k from 0.22 to 0.6 practically does not affect the link distributions $P(\kappa)$ changing the value of μ only by 10% (see Fig. 5). This shows that the correlations inside the directed network generated by the map [Eq. (2)] play a very important role.

E. Global contraction

As discussed above a nontrivial decay of the PageRank p_j in our Ulam network appears due to a dissipative nature of the map [Eq. (2)]. Indeed, since $\eta<1$ there is a global contraction of the phase space area by a factor $\Gamma_c=\eta^T$ after T iterations of the map (after its period). Such a property is very natural for the continuous map but it is more difficult to see its signature from the matrix form of the Perron-Frobenius operator after the introduction of discreteness of the phase space.

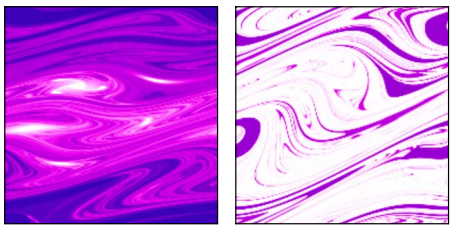

FIG. 10. (Color online) Same as Fig. 1 for the set $T10$ at $\alpha=0.99$, $N=3.6\times10^5$ at $k=0.22$ (left) and $k=0.6$ (right); PAR ξ are the same as in the inset of Fig. 9.

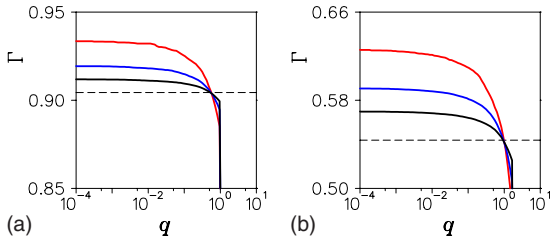

FIG. 11. (Color online) Dependence of the network contraction factor Γ on the level q of probability distribution over the network nodes (see text). Left panel shows data for the set $T10$ at $k=0.22$, right panel shows data for the set $T20$ at $k=0.3$ for the Ulam network of map [Eq. (2)]. The size of the network is $N=10^4,4\times10^4,16\times10^4$ (curves from top to bottom at $q=0.01$). The dashed line (in left/right panels) shows the contraction $\Gamma_c=\eta^T$ of the continuous map [Eq. (2)] corresponding to the network with $N=\infty$ and to values η and T for the sets $T10$ and $T20$, respectively.

Nevertheless this contraction can be extracted from the matrix $\mathbf{G}$ taken at $\alpha=1$. To extract it we apply $\mathbf{G}$ with $\alpha=1$ to a homogeneous vector $p_j^{(h)}=1/N$ getting the new vector $\bar{p}^{(h)}=\mathbf{G}p^{(h)}$ and count the number of nodes N_Γ where $\bar{p}^{(h)}>q/N$ and $0<q<1$ is some positive number characterizing the level of the distribution. Then the contraction of the network is defined as a fraction of such states: $\Gamma=N_\Gamma/N$.

The result of computation of the contraction factor for the Ulam network of the map [Eq. (2)] for the sets $T10$, $T20$ is shown in Fig. 11. The network contraction parameter Γ is independent of q in a large interval $10^{-4}\leq q\leq0.1$ and it converges to the contraction value Γ_c of a continuous map in the limit of large matrix size N.

We think that the Google matrix of WWW networks can be also characterized by a global contraction factor and it would be interesting to study its properties in more detail. However, this remains a task for future studies.

IV. SUMMARY

In summary, we demonstrated that the Perron-Frobenius operator built from a simple dissipative map with dynamical attractors generates a scale-free directed network with three properties similar to the WWW: compatible degree distributions (Figs. 4 and 5), compatible power-law PageRank decay (Fig. 6), its sensitivity for α near 1 and quasidegeneracy of eigenvalues near unity. The networks and their Google matrices are obtained on the basis of the Ulam method for coarse graining of the Perron-Frobenius operator and thus can be viewed as the Ulam networks or Ulam graphs. In this formulation the popular websites can be considered as dynamical fixed point attractors which help to generate global scale-free properties of the network. The PageRank of the system becomes delocalized for α smaller than a certain critical value, such a delocalization is linked to emergence of a strange attractor. Even for α very close to unity a moderate change in system parameters can drive the system to a strange attractor regime with a complete delocalization of the PageRank making the Google search inefficient. In view of a great importance of the Google search for WWW [3,7] and its new emerging applications [36] it may be rather useful to

study in more detail the properties of the Google matrix generated by simple dynamical maps. We will present our results for the real networks in a separate work [33].

At the present state the Google matrix of WWW is stable in respect to variations of α at typical values $\alpha=0.5-0.85$ (indications for that can be found e.g., in [9–11]). This is different from the cases $T10$ and $T20$ analyzed in our studies which show that there the PageRank vector becomes very flat in index for $\alpha<\alpha_c$ when the delocalization takes place. In this delocalized regime of the Ulam network the PageRank spreads over enormously large number of states growing to infinity with the increase in the matrix size N and hence the classification via the PageRank becomes inefficient. It is not excluded that, since WWW evolves with time, the WWW may become more sensitive to changes of α. Also the Google search can be applied to a large variety of other important networks (see, e.g., [11,36]) which may be more sensitive to various parameter variations.

Therefore, it is quite possible that the Ulam networks discussed here only partially simulate the properties of the WWW. For example our model shows certain deviations for distribution of links for small and very large number of links (see discussion above). However, the Ulam networks are easy to generate and at the same time they show a large variety of rich interesting properties. The parallels between the Ulam networks and the actual WWW can be instructive for deeper understanding of both. Therefore, we think that their further studies will give us better understanding of the Google matrix properties. The studies of the Ulam networks will also lead to a better understanding of intricate spectral properties of the Perron-Frobenius operators. The application of the thermodynamical formalism [37,38] to the spectra of such operators can help to understand their properties in a better way.

ACKNOWLEDGMENTS

We thank A. S. Pikovsky who pointed to us a link between our numerical construction procedure of the matrix **S** built from the discrete phase space cells and the Ulam method. One of us (D.L.S.) thanks A. S. Pikovsky for useful discussions and hospitality at the University Potsdam during the work on the revised version of this paper. We thank J. Palmeri for reading the text and corrections. We also thank an unknown referee who pointed to us Refs. [6,8–12] in the report for the initial short version of the paper.

APPENDIX

The Chirikov typical map [Eq. (2)] is studied here for the following random phases $\theta_t/2\pi$ for the set $T10$: 0.562 579, 0.279 666, 0.864 585, 0.654 365, 0.821 395, 0.981 145, 0.478 149, 0.834 115, 0.180 307, and 0.159 02, and for the set $T20$: 0.415 733 267 627, 0.310 795 551 489, 0.632 094 907 846, 0.749 488 203 411, 0.924 301 928 270, 0.635 937 571 045, 0.118 768 635 110, 0.647 524 548 037, 0.651 928 927 275, 0.952 312 529 146, 0.370 553 510 280, 0.810 837 257 644, 0.814 808 044 380, 0.834 758 628 241, 0.993 694 010 264, 0.702 057 578 688, 0.828 693 568 678, 0.855 421 638 697, 0.278 538 720 979, and 0.653 773 338 142. The numbers are ordered in the serpentine order for $t=1,2,\ldots,T$. After each T iterations the values of y are reduced inside the interval $(-\pi,\pi)$ corresponding to the periodic boundary conditions.

[1] See, e.g., http://www.worldwidewebsize.com/
[2] S. Brin and L. Page, Comput. Netw. ISDN Syst. **30**, 107 (1998).
[3] A. M. Langville and C. D. Meyer, *Google's PageRank and Beyond: The Science of Search Engine Rankings* (Princeton University Press, Princeton, 2006); D. Austin, AMS Feature Columns (2008) available at www.ams.org/featurecolumn/archive/pagerank.html
[4] I. P. Cornfeld, S. V. Fomin, and Y. G. Sinai, *Ergodic Theory* (Springer, New York, 1982).
[5] M. Brin and G. Stuck, *Introduction to Dynamical Systems* (Cambridge University Press, Cambridge, England, 2002).
[6] G. Osipenko, *Dynamical Systems, Graphs, and Algorithms* (Springer, Berlin, 2007).
[7] D. Donato, L. Laura, S. Leonardi, and S. Millozzi, Eur. Phys. J. B **38**, 239 (2004); G. Pandurangan, P. Raghavan, and E. Upfal, Internet Math. **3**, 1 (2005).
[8] P. Boldi, M. Santini, and S. Vigna, in *Proceedings of the 14th International Conference on World Wide Web*, edited A. Ellis and T. Hagino (ACM Press, New York, 2005), p. 557; S. Vigna, in *Proceedings of the 14th International Conference on World Wide Web*, edited A. Ellis and T. Hagino (ACM Press, New York, 2005), p. 976.
[9] K. Avrachenkov and D. Lebedev, Internet Math. **3**, 207 (2006).
[10] K. Avrachenkov, N. Litvak, and K. S. Pham, in *Algorithms and Models for the Web-Graph: Fifth International Workshop, WAW 2007 San Diego, CA, Proceedings*, edited by A. Bonato and F. R. K. Chung (Springer-Verlag, Berlin) Lect. Notes Comput. Sci. **4863**, 16 (2007).
[11] *Algorithms and Models for the Web-Graph: Sixth International Workshop, WAW 2009 Barcelona, Proceedings, Springer-Verlag, Berlin, Lecture Notes Computer Sci.* Vol. 5427, edited by K. Avrachenkov, D. Donato, and N. Litvak (Springer, Berlin, 2009).
[12] S. Serra-Capizzano, SIAM J. Matrix Anal. Appl. **27**, 305 (2005).
[13] M. L. Mehta, *Random Matrices*, 3rd ed. (Academic Press, New York, 2004).
[14] P. W. Anderson, Phys. Rev. **109**, 1492 (1958); P. A. Lee and T. V. Ramakrishnan, Rev. Mod. Phys. **57**, 287 (1985).
[15] O. Giraud, B. Georgeot, and D. L. Shepelyansky, Phys. Rev. E **72**, 036203 (2005).
[16] R. Berkovits, Eur. Phys. J. Spec. Top. **161**, 259 (2008).
[17] O. Giraud, B. Georgeot, and D. L. Shepelyansky, Phys. Rev. E **80**, 026107 (2009).

[18] B. V. Chirikov, *Research Concerning the Theory of Nonlinear Resonance and Stochasticity: Preprint No. 267* (Institute of Nuclear Physics, Novosibirsk, 1969) ([translation: CERN Trans. 71-40, Geneva (1971)]).
[19] K. M. Frahm and D. L. Shepelyansky, Phys. Rev. E **80**, 016210 (2009).
[20] S. M. Ulam, *A Collection of Mathematical Problems*, Interscience Tracs in Pure and Applied Mathematics, Vol. 8 (Interscience, New York, 1960), p. 73.
[21] T.-Y. Li, J. Approx. Theory **17**, 177 (1976).
[22] Z. Kovács and T. Tél, Phys. Rev. A **40**, 4641 (1989).
[23] Z. Kaufmann, H. Lustfeld, and J. Bene, Phys. Rev. E **53**, 1416 (1996).
[24] G. Froyland, R. Murray, and D. Terhesiu, Phys. Rev. E **76**, 036702 (2007).
[25] J. Ding and A. Zhou, Physica D **92**, 61 (1996).
[26] M. Blank, G. Keller, and C. Liverani, Nonlinearity **15**, 1905 (2002).
[27] D. Terhesiu and G. Froyland, Nonlinearity **21**, 1953 (2008).
[28] G. Froyland, S. Lloyd, and A. Quas, Ergod. Theory Dyn. Syst. **1**, 1 (2008).
[29] G. Froyland, in *Nonlinear Dynamics and Statistics: Proceedings, Newton Institute, Cambridge (1998)*, edited by A. Mees (Birkhäuser Verlag AG, Berlin, 2001), p. 283.
[30] D. Donato, L. Laura, S. Leonardi, and S. Millozzi, Comput. Sci. Eng. **6**, 84 (2004).
[31] B. V. Chirikov, Phys. Rep. **52**, 263 (1979).
[32] E. Ott, *Chaos in Dynamical Systems* (Cambridge University Press, Cambridge, 1993).
[33] B. Georgeot, O. Giraud, and D. L. Shepelyansky, e-print arXiv:1002.3342.
[34] S. Nonnenmacher and M. Zworski, Commun. Math. Phys. **269**, 311 (2007).
[35] D. L. Shepelyansky, Phys. Rev. E **77**, 015202(R) (2008).
[36] P. Chen, H. Xie, S. Maslov, and S. Redner, J. Informetrics **1**, 8 (2007).
[37] D. Ruelle, *Thermodynamical Formalism* (Cambridge University Press, Cambridge, England, 2004).
[38] R. Artuso, E. Aurell, and P. Cvitanović, Nonlinearity **3**, 325 (1990); **3**, 361 (1990).