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*

(Dated: May 26, 2009; Revised: August 20, 2009)

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 creates the directed Ulam networks with approximate scale-free scaling and characteristics being rather similar to those of the World Wide Web. The simple dynamical attractors play here the role of popular web sites with a strong concentration of PageRank. A variation of 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.

PACS numbers: 05.45.-a, 89.20.Hh, 05.45.Ac

I INTRODUCTION

The World Wide Web (WWW) continues its striking expansion going beyond 10^{11} web pages. 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 [1], forms the basis of the Google search engine used by the majority of internautes in everyday life. The PRA is based on the construction of the Google matrix which can be written as (see e.g. [2] for details):

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

Here the matrix $\mathbf{S}$ is constructed from the adjacency matrix $\mathbf{A}$ of directed network links between N nodes so that $S_{ij} = A_{ij}/\sum_k A_{kj}$ and the elements of columns with only zero elements are replaced by $1/N$. The second term in r.h.s. of (1) describes a finite probability $1-\alpha$ for WWW surfer to jump at random to any node so that the matrix elements $E_{ij} = 1$. This term allows to stabilize the convergence of PRA introducing a gap between the maximal eigenvalue $\lambda = 1$ and other eigenvalues λ_i. Usually the Google search uses the value $\alpha = 0.85$ [2]. By the construction $\sum_i G_{ij} = 1$ so that the asymmetric matrix $\mathbf{G}$ has a left eigenvector being a homogeneous constant for $\lambda = 1$. The right eigenvector at $\lambda = 1$ is the PageRank vector with positive elements p_j and $\sum_j p_j = 1$. All WWW nodes can be ordered by decreasing p_j so that the PageRank plays a primary role in the ordering of websites and information retrieval. The classification of nodes in the decreasing order of p_j values is used by the Google search to classify importance of web nodes. 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 the construction the operator $\mathbf{G}$ belongs to the class of Perron-Frobenius operators [2]. Such type of operators naturally appear in the ergodic theory [3] and in the description of dynamical systems with Hamiltonian or dissipative dynamics [4, 5].

The studies of properties of $\mathbf{G}$ are usually done only for the PageRank vector which can be find efficiently by the PRA due to a relatively small average number of links in WWW. At present Google succeeds to operate with PageRank vectors of size of the whole WWW being of the order of 10^{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 $\beta \approx 0.9$ [2, 6]. The studies of PageRank properties are now very active in the computer science community being presented in a number of interesting publications (see e.g. [7, 8, 9] and an overview of the field in [10]).

While the properties of the PageRank 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 [11] which finds applications for a description of spectra in complex many-body quantum systems and the Anderson localization which is an important physical phenomenon for an electron transport in disordered systems (see e.g. [12]). A transition from localized to delocalized eigenstates also can take place in networks of small world type (see [13, 14]). However, in the physical systems considered in [11, 12, 13, 14] all matrices are Hermitian with real eigenvalues while the Perron-Frobenius matrices have generally complex eigenvalues.

A first attempt to analyze the properties of right eigenvectors ψ_i ($\mathbf{G}\psi_i = \lambda_i \psi_i$) and complex eigenvalues λ_i was done recently in [15]. The Google matrix was constructed from a directed network generated by the Albert-Barabasi model and the WWW University net-

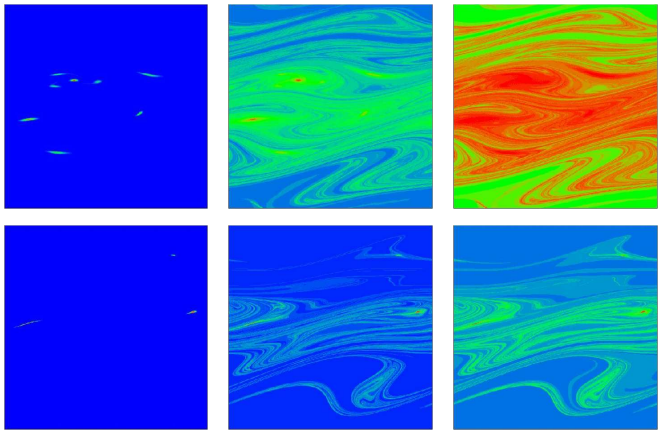

FIG. 1: (Color online) PageRank p_j for the Google matrix generated by the Chirikov typical map (2) at $T = 10$, $k = 0.22$, $\eta = 0.99$ (set $T10$, top row) and $T = 20$, $k = 0.3$, $\eta = 0.97$ (set $T20$, bottom row) with $\alpha = 1, 0.95, 0.85$ (left to right). The phase space region $0 \leq x < 2\pi; -\pi \leq p < \pi$ is divided on $N = 3.6 \cdot 10^5$ cells; p_j is zero for blue and maximal for red.

works with randomization of links. The Google matrix was considered mainly for the value $\alpha = 0.85$. It was shown that at certain conditions a delocalization phase emerges for the PageRank and states with complex λ. In spite of a number of interesting results found in [15] a weak feature of models used there is a significant gap between $\lambda = 1$ of PageRank vector and $|\lambda_i| \leq 0.4$ of other vectors. We note that according to [15] 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 was not sensitive to a variation of α in the interval $0.85 \leq \alpha \leq 1$. Hence, the PageRank vector also was not very sensitive to α while for real WWW it is know that p_j is rather sensitive to α due to existence of $|\lambda_i|$ close to 1 [2, 15]. Thus the results obtained in [15] 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 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 have 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 these models has rich and nontrivial features of general importance like 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 [16] with dissipation, the properties of this simple model has been analyzed in detail in a recent work [17]. 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 captures also other specific properties of WWW Google matrices. 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 on 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 $\sum_i S_{ij} = 1$). Thus $\mathbf{S}$ gives a coarse-grained approximation of the Perron-Frobenius operator for the dynamical map. The Google matrix $\mathbf{G}$ of size N is constructed from $\mathbf{S}$ according to Eq. (1). We use a sufficiently large values of N_c so that the properties of $\mathbf{G}$ become not sensitive to N_c.

Such a discrete approximation of the Perron-Frobenius operator is known in dynamical systems as the Ulam method [18]. 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 [19]. Various types of more generic one-dimensional maps have been studied in [20, 21, 22]. Further mathematical results have been reported in [23, 24, 25, 26] with extensions and prove of convergence for hyperbolic maps in higher dimensions. However, the studies of more generic two-dimensional maps remain rather restricted (see e.g. [27]) and non-systematic. In principle the construction of directed graphs on the basis of dynamical systems is a known mathematical approach (see e.g. [5]) but the spectral properties of the Google matrix built on such graphs were not studied till now.

In this paper we show that the Ulam method applied to two-dimensional dissipative dynamical maps generates a new 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 Section II we give the description 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 the map and the network are described here; in Section III the properties of the eigenvalues and eigenstates of the Google matrix are analyzed in detail including the delocalization transition for the PageRank, fractal Weyl law and the global contraction properties; the summary of the results is presented in Section IV.

II ULAM NETWORKS OF DYNAMICAL MAPS

Chirikov typical map

To construct an Ulam network and a generated by it Google matrix we use a dynamical two-dimensional dissipative map. The dynamical system is described by the Chirikov typical map introduced in 1969 for a description of continuous chaotic flows [16]:

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

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 [17]. 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 [3, 4, 28, 29]).

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 $\eta = 1$ [17]. 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. [29]) 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 Fig. 2 and Fig. 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.

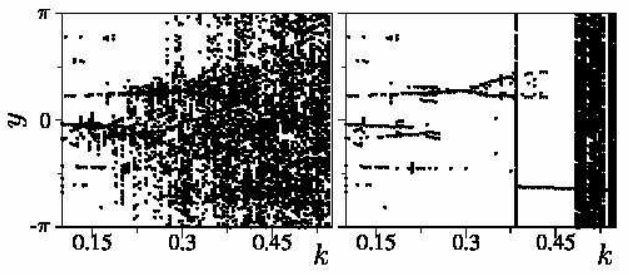

FIG. 2: Bifurcation diagram showing values of y vs. map parameter k for the set $T10$ of the Chirikov typical map (2). The values of y, obtained from 10 trajectories with initial random positions in the phase space region, are shown for integer moments of time $100 < t/T \le 110$ (left panel) and $10^4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

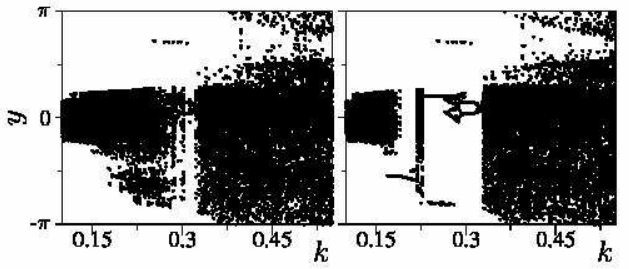

FIG. 3: Same as in Fig. 2 for the set $T20$.

This means that during a long time a trajectory moves between few simple attractors (which are clearly seen in Fig. 1 in the left column) before a final convergence is reached.

Network construction and distribution of links

The Ulam network for the Chirikov typical map (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 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 $\sum_i S_{ij} = 1$. Such $\mathbf{S}$ gives a coarse-grained approximation of the Perron-Frobenius operator for the map (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 of N_c in the interval $10^3 \le N_c \le 10^5$. Since the cell size is very small it is

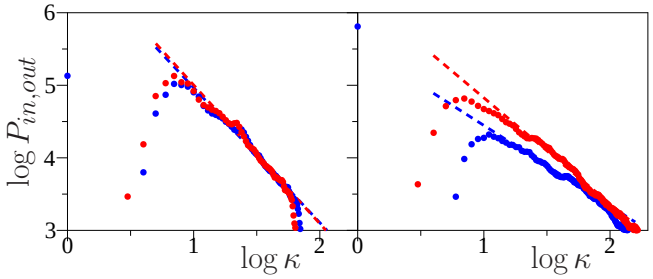

FIG. 4: (Color online) Differential distribution of number of nodes with *ingoing* $P_{in}(\kappa)$ (blue) and *outgoing* $P_{out}(\kappa)$ (red) links κ for sets $T10$ (left) and $T20$ (right). The straight dashed lines give the algebraic fit $P(\kappa) \sim \kappa^{-\mu}$ with the exponent $\mu = 1.86, 1.11$ ($T10, T20$) for *ingoing* and $\mu = 1.91, 1.46$ ($T10, T20$) *outgoing* links. Here $N = 1.44 \cdot 10^6$ and $P(\kappa)$ gives a number of nodes at a given integer number of links κ for this matrix size. Blue point at $\kappa = 0$ shows that in the whole matrix there is a significant number of nodes with zero *ingoing* links.

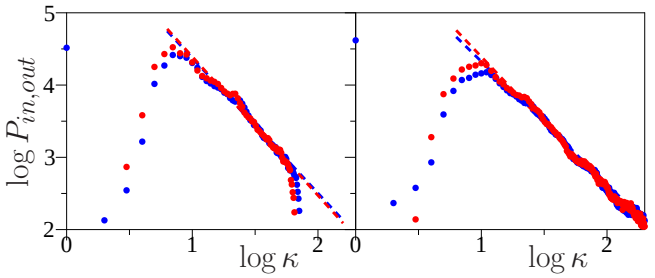

FIG. 5: (Color online) Same as in Fig. 3 for the set $T10$ at $k = 0.22$ (left) (same as Fig. 4 left) and $k = 0.6$ (right) and $N = 3.6 \cdot 10^5$. The fit gives the exponent $\mu = 1.87, 1.92$ for *ingoing* (blue), *outgoing* (red) links at $k = 0.22$ (left) and $\mu = 1.70, 1.83$ for *ingoing* (blue), *outgoing* (red) links at $k = 0.6$ (right).

unimportant in what way N_c trajectories are distributed inside the cell. Up to statistical fluctuations, the values of S_{ij} remains the same for homogeneous or random distribution of N_c trajectories inside a cell.

Up to $N = 22500$ we used exact diagonalization of $\mathbf{G}$ to determine all eigenvalues λ_i and right eigenvectors ψ_i, for larger N up to $N = 1.44 \cdot 10^6$ we used the PRA to determine the PageRank vector. The majority of data are presented for two typical sets $T10, T20$ of parameters of the map (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 $t \sim 10^3$ 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 (2): effectively it introduces an additional noise which amplitude σ is approximately $\sigma \sim 2\pi/\sqrt{N}$. 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 [28] and the continuous map dynamics is bounded by the invariant Kolmogorov-Arnold-Moser (KAM) curves. However, the discreteness of phase space allows to jump from one cell to another and thus to jump from one curve to another. This leads to a diffusion in y direction and appearance of a homogeneous ergodic state at $\lambda = 1$. A direct analysis also shows that at any finite cell size the operator $\mathbf{S}$ has a homogeneous ergodic state with $\lambda = 1$, we also checked this via numerical diagonalization of matrix sizes $N \approx 20000$. 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 r.h.s. of (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 $\sigma \sim 2\pi/\sqrt{N}$ is an effective amplitude of noise introduced by cell discreetness.

Even if this σ-noise leads to a drastic change of 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 discreetness 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. 4,5 we show the distributions of ingoing $P_{in}(\kappa)$ and outgoing $P_{out}(\kappa)$ links κ in the Ulam network presented by $\mathbf{S}$ matrix generated by the map (2) as described above. These distributions are satisfactory described by a scale-free algebraic decay $P \sim 1/\kappa^{\mu}$ with $\mu \approx 1.86, 1.11$ for ingoing and $1.91, 1.46$ outgoing links at $T10, T20$ respectively and a typical number of links per node $\kappa \sim 10$ (see Fig. 4 and Fig. 5). Such values are compatible with the WWW data of scale-free type where $\mu \approx 2.1, 2.7$ for ingoing, outgoing links [2, 6]. 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 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 $\kappa \approx 50$. 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 $N \sim 10^5 - 10^6$.

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 $\psi_i^+ = const$. 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.

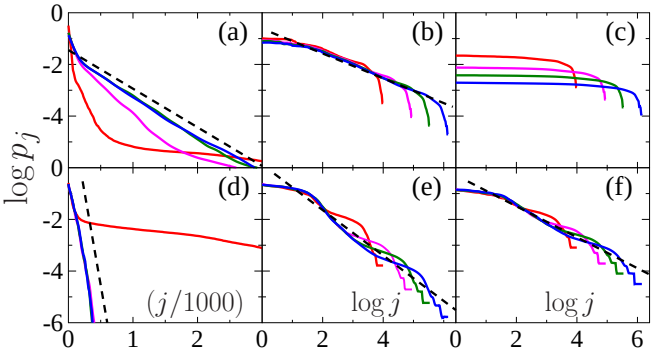

FIG. 6: (Color online) PageRank distribution p_j for $N = 10^4$, $9 \cdot 10^4$, $3.6 \cdot 10^5$ and $1.44 \cdot 10^6$ shown by red, magenta, green and blue curves, the dashed straight lines show fits $p_j \sim 1/j^\beta$ with β: 0.48 (b), 0.88 (e), 0.60 (f). Dashed lines in panels (a),(d) show an exponential 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),(d) the curves at large N become superimposed. Here and below logarithms are decimal.

III PROPERTIES OF EIGENVALUES AND EIGENSTATES

Delocalization transition for PageRank with α

The variation of 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 of α the PageRank starts to spread over a strange attractor set. The properties of strange attractors in dynamical dissipative systems are described in [29]. In the map (2) the strange attractor appears at larger values of k (namely $k > 0.5$ for $T10$, $k > 0.34$ for $T20$, see Figs. 2, 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 Fig. 6a,d). Such an exponential decay results from the Fokker-Planck description of map (2) in the presence of σ noise term which gives diffusive transitions on nearby cells. For $\alpha < 1$ random surfer transitions 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 (Fig. 6b,e,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 [2]. However, β decreases with the decrease of 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. 6c). For $T20$ set $p_j \sim \psi_{i=1}(j)$ remains localized at $\alpha = 0.85$ so that a PArticipation Ratio (PAR) $\xi = \sum_j (|\psi_i(j)|^2)^2 / \sum_j (|\psi_i(j)|^4$ for the PageRank remains finite at large N. We use this definion of PAR ξ for all eigenvectors $\psi_i(j)$.

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. 7a). 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 Fig. 7c,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 of 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. 7b). We attribute this to the fact

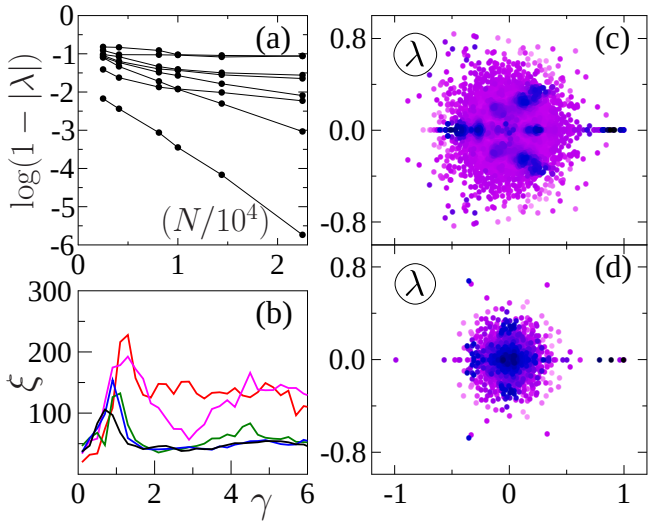

FIG. 7: (Color online) (a) Dependence of gap $1-|\lambda|$ on Google matrix size N for few eigenstates with $|\lambda|$ most close to 1, set $T10$, $\alpha = 1$; (b) dependence of PAR ξ on $\gamma = -2\ln|\lambda|$ for $N = 2500$, 5625, 8100, 10^4, 14400 for set $T10$, $\alpha = 1$ (curves from top to bottom: 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 $\alpha = 1$, $N = 1.44 \cdot 10^4$); (d) same as (c) but for set $T20$.

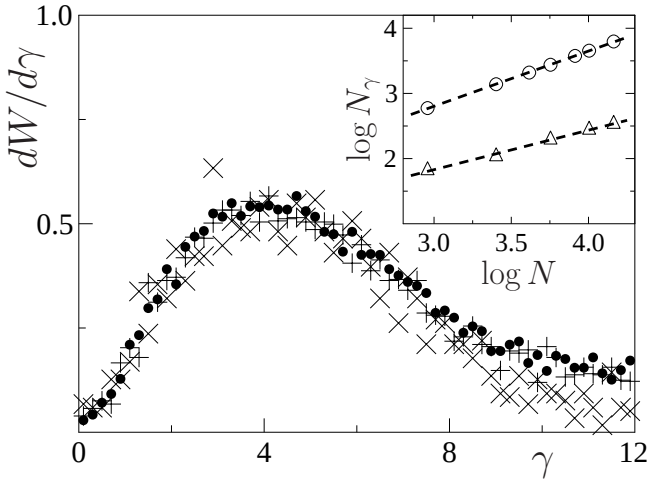

FIG. 8: Probability distribution $dW(\gamma)/d\gamma$ for set $T10$, $\alpha = 1$ at $N = 2.5 \cdot 10^3(\times)$, $10^4(+)$, $1.44 \cdot 10^4$ (dots); $W(\gamma)$ is normalized by the number of states $N_\gamma = 0.55N^{0.85}$ with $\gamma < 6$. Inset: dependence of number of states N_γ with $\gamma < \gamma_b$ 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.

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 obtained in [15] 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 [15]): 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. 7b. 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 $0 < |\lambda_i| < 1$. In agreement with the known theorems [2] 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 $(1-\alpha)$ according to [2]). 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.

Fractal Weyl law for Google matrix

Another interesting characteristic of $\mathbf{G}$ is the density distribution $dW(\gamma)/d\gamma$ over γ. 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 rather interesting 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 non-unitary matrices with fractal eigenstates (see e.g. [30, 31] and Refs. therein). According to this law the exponent ν is $\nu = d-1$ where d is the fractal dimension of the system. Approximately we have $d - 1 \approx 1 - \gamma_c/(Th)$ [29, 31] that gives $\nu = 0.88, 0.72$ for the sets $T10, T20$ 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 [30, 31] and Refs. 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.

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 the 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 $T10, T20$. 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 Fig. 7c,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 [31] 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 transition from localized to delocalized PageRank by increasing k in the map (2) (see Fig. 9 inset and Fig. 10). This transition in k takes place approximately at $k \approx 0.55$ when fixed point attractors merge into a strange attractor (see the bifurcation diagram in Fig. 2). A peak in ξ around $k \approx 0.38$ is related to birth and disappearance of a strange attractor in a narrow interval of k at $k \approx 0.38$. At the same time an increase of 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 (2) play a very important role.

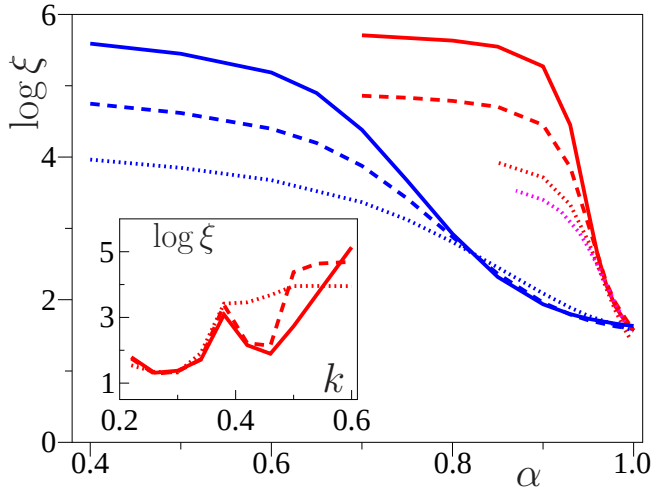

FIG. 9: (Color online) Dependence of PageRank ξ on α for set $T10$ at $N = 5625$ (dotted magenta), $1.44 \cdot 10^4$ (dotted red), $9 \cdot 10^4$ (dashed red), $6.4 \cdot 10^5$ (full red) and for $T20$ at $N = 1.44 \cdot 10^4$ (dotted blue), $9 \cdot 10^4$ (dashed blue), $6.4 \cdot 10^5$ (full blue). Inset shows dependence of ξ on k for set $T10$ at $\alpha = 0.99$ with $N = 1.44 \cdot 10^4$ (dotted red), $9 \cdot 10^4$ (dashed red), $3.6 \cdot 10^5$ (full red).

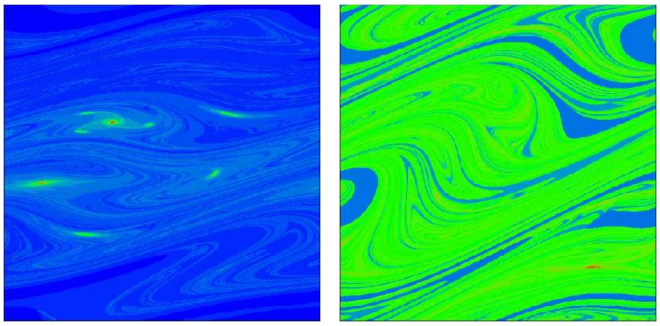

FIG. 10: (Color online) Same as Fig. 1 for the set $T10$ at $\alpha = 0.99$, $N = 3.6 \cdot 10^5$ at $k = 0.22$ (left) and $k = 0.6$ (right); PAR ξ are the same as in the inset of Fig. 9.

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 (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.

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 map (2) for the sets $T10$, $T20$ is shown in Fig. 11. The network contraction parameter Γ is independent of q in a large interval $10^{-4} \leq q \leq 0.1$ 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.

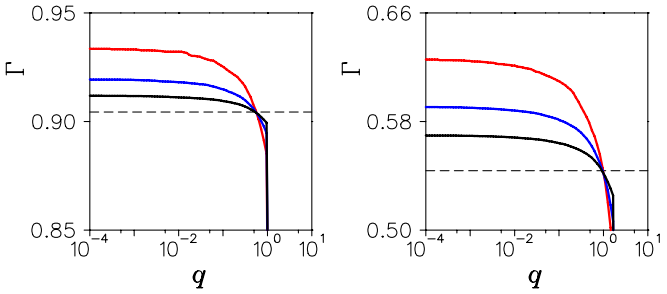

FIG. 11: (Color online) Dependence of the network contraction factor Γ on the level q of probability distribution over the network nodes (see text). Left panel shows data for the set $T10$ at $k = 0.22$, right panel shows data for the set $T20$ at $k = 0.3$ for the Ulam network of map (2). The size of the network is $N = 10^4, 4 \cdot 10^4, 16 \cdot 10^4$ (curves from top to bottom at $q = 0.01$). The dashed curves show the contraction $\Gamma_c = \eta^T$ of the continuous map (2) corresponding to the network with $N = \infty$.

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 properties being rather similar to the WWW. 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. The Google matrix of such Ulam networks reproduces many properties of real networks with an algebraic decay of the PageRank and quasi-degeneracy of eigenvalues near unity for the Google parameter $\alpha = 1$. In this formulation the popular websites correspond to 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 of 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 [2, 6] and its new emerging applications [32] it may be rather useful to study in more detail the properties of the Google matrix generated by simple dynamical maps.

Of course, it is quite possible that at the present state the Google matrix of WWW is more stable in respect to variation of α (indications for that can be found e.g. in [8, 9, 10]). However, WWW evolves with time and 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. [10, 32]) which may be more sensitive to various parameter variations. It is quite possible that the Ulam networks discussed here only partially simulate the properties of the WWW. 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 [33, 34] to the spectra of such operators can help to understand their properties in a better way.

ACKNOWLEDGEMENTS

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 (DLS) thanks A.S.Pikovsky for useful discussions and hospitality at the Univ. Potsdam during the work on the revised version of this paper. We also thank an unknown referee B who pointed to us Refs. [5, 7, 8, 9, 10] in the report for the initial short version of the paper.

[1] S. Brin and L. Page, Computer Networks and ISDN Systems **33**, 107 (1998).

[2] 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

[3] I.P. Cornfeld, S.V. Fomin, and Y. G. Sinai, *Ergodic theory*, Springer, N.Y. (1982).

[4] M. Brin and G. Stuck, *Introduction to dynamical systems*, Cambridge Univ. Press, Cambridge, UK (2002).

[5] G. Osipenko, *Dynamical systems, graphs, and algorithms*, Springer, Berlin (2007).

[6] 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).

[7] P. Boldi, M. Santini, and S. Vigna, in *Proceedings of the 14th international conference on World Wide Web*, A. Ellis and T. Hagino (Eds.), ACM Press, New York p.557 (2005); S. Vigna, **ibid.** p.976.

[8] K. Avrachenkov and D. Lebedev, Internet Mathematics **3**, 207 (2006).

[9] K. Avrachenkov, N. Litvak, and K.S. Pham, in *Algorithms and Models for the Web-Graph: 5th International Workshop, WAW 2007 San Diego, CA, Proceedings*, A. Bonato and F.R.K. Chung (Eds.), Springer-Verlag, Berlin, Lecture Notes Computer Sci. **4863**, 16 (2007)

[10] K. Avrachenkov, D. Donato and N. Litvak (Eds.), *Algorithms and Models for the Web-Graph: 6th Interna-*

tional Workshop, WAW 2009 Barcelona, Proceedings, Springer-Verlag, Berlin, Lecture Notes Computer Sci. **5427**, Springer, Berlin (2009).
[11] M.L. Mehta, *Random matricies*, Academic Press Inc., N.Y., 3d Ed. (2004).
[12] P. W. Anderson, Phys. Rev. **109**, 1492 (1958); P.A. Lee and T.V. Ramakrishnan, Rev. Mod. Phys. **57**, 287 (1985).
[13] O. Giraud, B. Georgeot and D.L. Shepelyansky, Phys. Rev. E **72**, 036203 (2005).
[14] R. Berkovits, Eur. Phys. J. Special Topics **161**, 259 (2008).
[15] O. Giraud, B. Georgeot and D. L. Shepelyansky, Phys. Rev. E **80**, 026107 (2009).
[16] B.V. Chirikov, *Research concerning the theory of nonlinear resonance and stochasticity*, Preprint N 267, Institute of Nuclear Physics, Novosibirsk (1969) [translation: CERN Trans. 71 - 40, Geneva (1971)].
[17] K.M. Frahm and D.L. Shepelyansky, Phys. Rev. E **80**, 016210 (2009).
[18] S.M. Ulam, *A Collection of mathematical problems*, Vol. 8 of *Interscience tracs in pure and applied mathematics*, Interscience, New York, p. 73 (1960).
[19] T.-Y. Li, J. Approx. Theory **17**, 177 (1976).
[20] Z. Kovács and T. Tél, Phys. Rev. A **40**, 4641 (1989).
[21] Z. Kaufmann, H. Lustfeld, and J. Bene, Phys. Rev. E **53**, 1416 (1996).
[22] G. Froyland, R. Murray, and D. Terhesiu, Phys. Rev. E **76**, 036702 (2007).
[23] J. Ding and A. Zhou, Physica D **92**, 61 (1996).
[24] M. Blank, G. Keller, and C. Liverani, Nonlinearity **15**, 1905 (2002).
[25] D. Terhesiu and G. Froyland, Nonlinearity **21**, 1953 (2008).
[26] G. Froyland, S. Lloyd, and A. Quas, Ergod. Th. Dynam. Sys. **1**, 1 (2008).
[27] G. Froyland, *Extracting dynamical behaviour via Markov models*, in A. Mees (Ed) *Nonlinear Dynamics and Statistics: Proceedings, Newton Institute, Cambridge (1998)*, p.283 Birkhäuser Verlag AG, Berlin (2001).
[28] B.V. Chirikov, Phys. Rep. **52**, 263 (1979).
[29] E. Ott, *Chaos in Dynamical Systems*, Cambridge Univ. Press, Cambridge (1993).
[30] S. Nonnenmacher and M. Zworski, Comm. Math. Phys. **269**, 311 (2007).
[31] D. L. Shepelyansky, Phys. Rev. E **77**, 015202(R) (2008).
[32] P. Chen, H. Xie, S. Maslov and S. Redner, J. Informetrics **1**, 8 (2007).
[33] D. Ruelle, *Thermodynamical formalism*, Cambridge Univ. Press, Cambridge, UK (2004).
[34] R. Artuso, E. Aurell, and P. Cvitanović, Nonlineariy **3**, 345 (1990); **ibid/ 3**, 361 (1990).

APPENDIX

The Chirikov typical map (2) is studied here for the following random phases $\theta_t/2\pi$ for the set $T10$:
0.562579, 0.279666, 0.864585, 0.654365, 0.821395, 0.981145, 0.478149, 0.834115, 0.180307, 0.15902
and for the set $T20$:
0.415733267627, 0.310795551489, 0.632094907846, 0.749488203411, 0.924301928270, 0.635937571045, 0.118768635110, 0.647524548037, 0.651928927275, 0.952312529146, 0.370553510280, 0.810837257644, 0.814808044380, 0.834758628241, 0.993694010264, 0.702057578688, 0.828693568678, 0.855421638697, 0.278538720979, 0.653773338142.

The numbers are ordered in the serpentine order for $t = 1, 2, ...T$.

After each T iterations the values of y are reduced inside the interval $(-\pi, \pi)$ corresponding to the periodic boundary conditions.