Google matrix and dynamical attractors

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

[1] *Université de Toulouse, UPS, Laboratoire de Physique Théorique (IRSAMC), F-31062 Toulouse, France*
[2] *CNRS, LPT (IRSAMC), F-31062 Toulouse, France*
[3] *Budker Institute of Nuclear Physics, 630090 Novosibirsk, Russia*

(Dated: May 26, 2009)

We study the properties of the Google matrix generated by a coarse-grained Perron-Frobenius operator of the Chirikov typical map with dissipation. This simple dynamical model creates scale-free directed networks with 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 system parameters drives the system 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

The World Wide Web (WWW) continues its striking expansion going beyond 10^{11} web pages. Information retrieval from such an enormous data base 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 at 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 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. By the construction the operator $\mathbf{G}$ belongs to the class of Perron-Frobenius operators [2].

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. 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, 3]. 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 [4]. The Google matrix was constructed from a directed network generated by the Albert-Barabasi model and the WWW University networks with randomization of links. 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 [4] a weak feature of models used there is a significant gap between $\lambda = 1$ of PageRank vector and λ_i of other vectors. Thus the PageRank was not very sensitive to α while for real WWW (without randomization of links) it is know that p_j is rather sensitive to α due to existence of λ_i close to 1 [2, 4]. In this work we use another approach and construct the Google matrix from the Perron-Frobenius operator generated by a dynamical system described by the Chirikov typical map [5] with dissipation. This model has many λ_i close to 1 and the PageRank becomes sensitive to α (see Fig. 1). We find that it captures also other specific properties of real Google matrix.

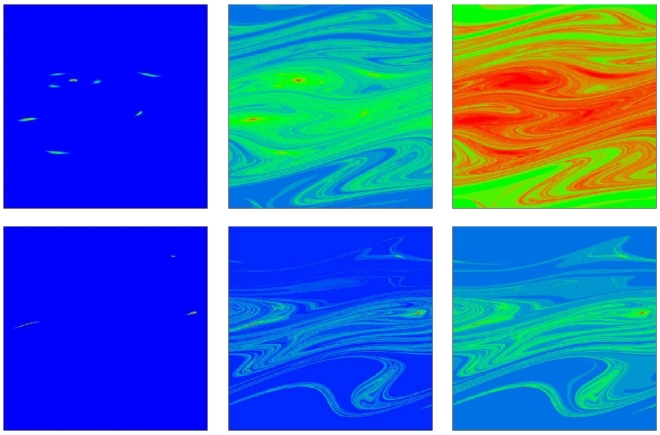

FIG. 1: (Color online) PageRank p_j for the Google matrix generated by the Chirikov typical map (2) at $T = 10$, $k = 0.22$, $\eta = 0.99$ (set $T10$, top row) and $T = 20$, $k = 0.3$, $\eta = 0.97$ (set $T20$, bottom row) with $\alpha = 1, 0.95, 0.85$ (left to right). The phase space region $0 \le x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

The dynamical system is described by the Chirikov

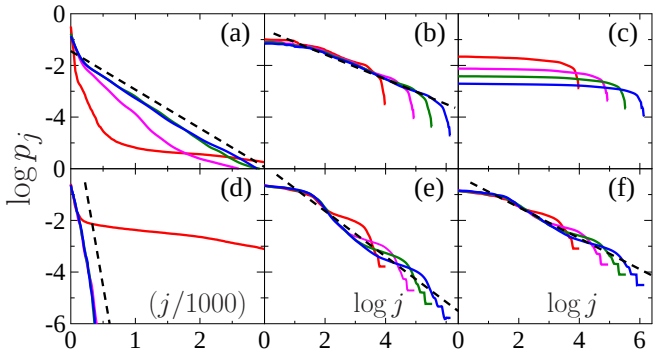

FIG. 2: (Color online) PageRank distribution p_j for $N = 10^4$, $9 \cdot 10^4$, $3.6 \cdot 10^5$ and $1.44 \cdot 10^6$ shown by red, magenta, green and blue curves, the dashed straight lines show fits $p_j \sim 1/j^\beta$ with β: 0.48 (b), 0.88 (e), 0.60 (f). Dashed lines in panels (a),(d) show an exponential Boltzmann decay (see text, lines are shifted in j for clarity). Other parameters 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.

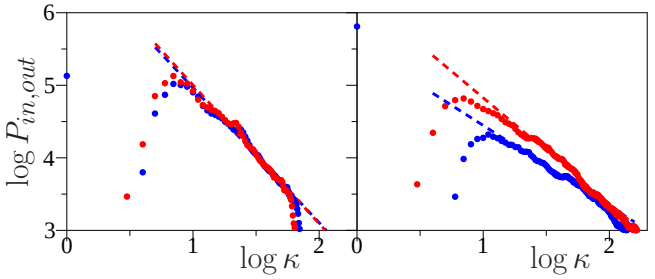

FIG. 3: (Color online) Differential distribution of number of nodes with *ingoing* $P_{in}(\kappa)$ (blue) and *outgoing* $P_{out}(\kappa)$ (red) links κ for sets $T10$ (left) and $T20$ (right). The straight dashed lines give the algebraic fit $P(\kappa) \sim \kappa^{-\mu}$ with the exponent $\mu = 1.86, 1.11$ ($T10, T20$) 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.

typical map introduced in 1969 for a description of continuous chaotic flows [5]:

$$\bar{p} = \eta p + k \sin(x + \theta_t) \ , \quad \bar{x} = x + \bar{p} \ . \qquad (2)$$

Here the bars note the new values of variables and $\theta_t = \theta_{t+T}$ are T random phases periodically repeated in time t. We consider the map in the region of Fig. 1 with the periodic boundary conditions. The parameter $0 < \eta \leq 1$ gives the global dissipation. The properties of the symplectic map at $\eta = 1$ have been studied recently in detail [6]. The dynamics is globally chaotic for $k > k_c \approx 2.5/T^{3/2}$ and the Kolmogorov-Sinai entropy is $h \approx 0.29 k^{2/3}$.

The Google matrix for the map (2) is built in the following way: the whole phase space region is divided into $N = N_x \times N_p$ cells ($N_x = N_p$), N_c trajectories are propagated from a cell j on T map iterations 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 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 $10^3 \leq N_c \leq 10^5$. A finite cell size corresponds physically to an addition of noise of amplitude $\sigma \sim 2\pi/\sqrt{N}$ in r.h.s. of (2). 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 (2) shown in Fig. 1 with numerically computed $h = 0.085, 0.108$ for $T10, T20$. 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. The distribution of ingoing $P_{in}(\kappa)$ and outgoing $P_{out}(\kappa)$ links κ in $\mathbf{S}$ is 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. 3 and Fig. 7 of Appendix). Such values are compatible with the WWW data of scale-free type where $\mu \approx 2.1, 2.7$ for ingoing, outgoing links [2, 3]. In our model a large number of links appears due to exponential stretching of one cell after T map iterations that gives $k \sim \exp(hT)$.

The variation of PageRank p_j with α is shown in Fig. 1 for two sets. At $\alpha = 1$ 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 [7]. 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. 8, 9 of Appendix) 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. 2. 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. 2a,d). Such an exponential decay results from the Fokker-Planck description of map (2) in the presence of σ noise term which

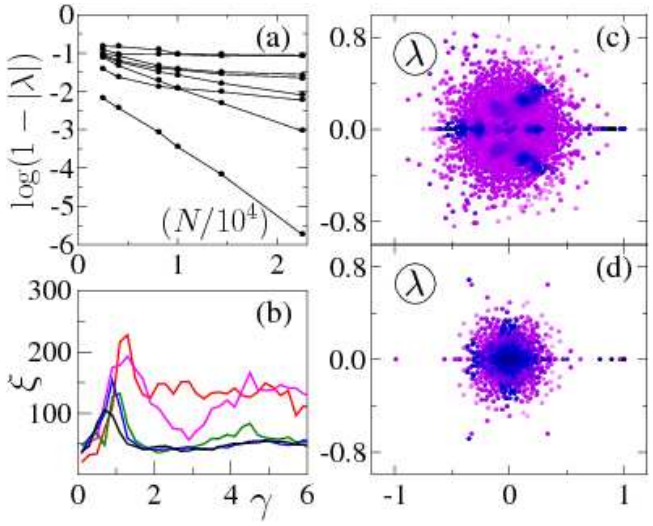

FIG. 4: (Color online) (a) Dependence of gap $1-|\lambda|$ on Google matrix size N for few eigenstates with $|\lambda|$ most close to 1, set $T10$, $\alpha = 1$; (b) dependence of PAR ξ on $\gamma = -2\ln|\lambda|$ for $N = 2500$, 5625, 8100, 10^4, 14400 for set $T10$, $\alpha = 1$ (curves from top to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

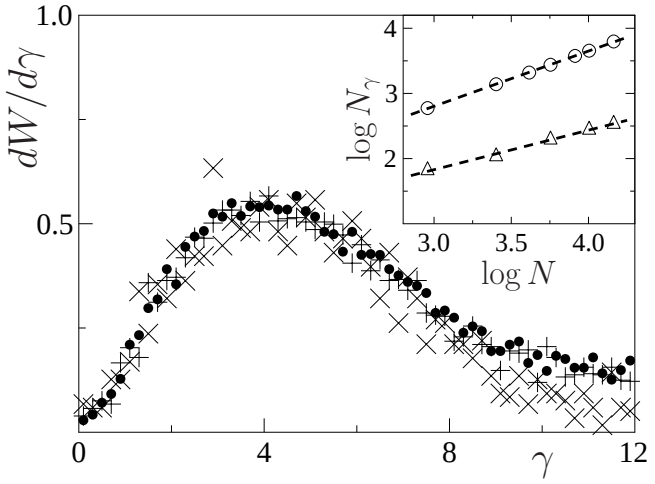

FIG. 5: Probability distribution $dW(\gamma)/d\gamma$ for set $T10$, $\alpha = 1$ at $N = 2.5 \cdot 10^3(\times)$, $10^4(+)$, $1.44 \cdot 10^4$ (dots); $W(\gamma)$ is normalized by the number of states $N_\gamma = 0.55N^{0.85}$ with $\gamma < 6$. Inset: dependence of number of states N_γ with $\gamma < \gamma_b$ 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.

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. 2b,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. 2c). 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.

To understand the origin of the delocalization transition in α we analyze in Fig. 4 the properties of all eigenvalues λ_i and eigenvectors ψ_i with their PAR ξ. Due to σ noise activation transitions take place between these fixed points leading to states with λ_i being exponentially close to $\lambda = 1$ (Fig. 4a). The distribution of λ_i in the complex plan is shown in Fig. 4c,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. 4b). 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 obtained in [4] 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 [4]): 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. 4b. Globally our data of Fig. 4 show that the diffusive modes at $|\lambda_i| < 1$ remain localized on a number of nodes $\xi \ll N$. 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.

Another interesting characteristic of $\mathbf{G}$ is the density distribution $dW(\gamma)/d\gamma$ over γ. The data presented in Fig. 5 show that it 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. 5 inset). We interpret this result on the basis of the fractal Weyl law established recently for matrices with fractal eigenstates (see e.g. [8] and Refs. therein). According to this law the exponent $\nu = d - 1$ where d is the fractal dimension of the system. Approximately we have $d - 1 \approx 1 - \gamma_c/(Th)$ [7, 8] 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. 5 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 dis-

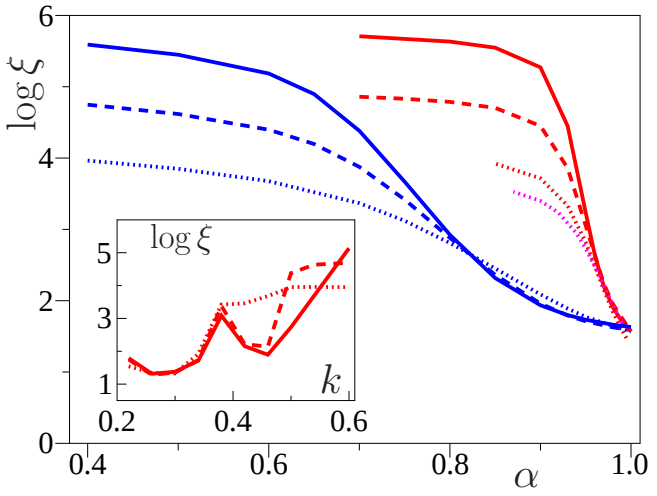

FIG. 6: (Color online) Dependence of PageRank ξ on α for set $T10$ at $N = 5625$ (dotted magenta), $1.44 \cdot 10^4$ (dotted red), $9 \cdot 10^4$ (dashed red), $6.4 \cdot 10^5$ (full red) and for $T20$ at $N = 1.44 \cdot 10^4$ (dotted blue), $9 \cdot 10^4$ (dashed blue), $6.4 \cdot 10^5$ (full blue). 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).

cuss the properties of these degenerate states with large $\xi \sim N$).

The dependence of PAR ξ of the PageRank on α and N is shown in Fig. 6. It allows to determine the critical value α_c below which the PageRank becomes delocalized showing ξ growing 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. 4c,d) and thus larger values of $1 - \alpha$ are required to get a significant contribution of these excited relaxation modes to the PageRank. Also the data of [8] 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 have 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. 6 inset and Fig. 10 of Appendix). 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 Figs. 8 of Appendix). 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 distributions $P(\kappa)$ changing the value of μ only by 10%. This shows that the correlations inside the directed network generated by the map (2) play a very important role.

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 Google matrix of this dynamical system 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, 3] and its new emerging applications [9] it may be rather useful to study in more detail the properties of the Google matrix generated by simple dynamical maps.

[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] 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).

[4] O. Giraud, B. Georgeot and D. L. Shepelyansky, arXiv:0903.5172[cs.IR] (2009).

[5] 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)].

[6] K.M. Frahm and D.L. Shepelyansky, arXiv:0905.1884[cond-mat] (2009).

[7] E. Ott, *Chaos in Dynamical Systems*, Cambridge Univ. Press, Cambridge (1993).

[8] D. L. Shepelyansky, Phys. Rev. E **77**, 015202(R) (2008).

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

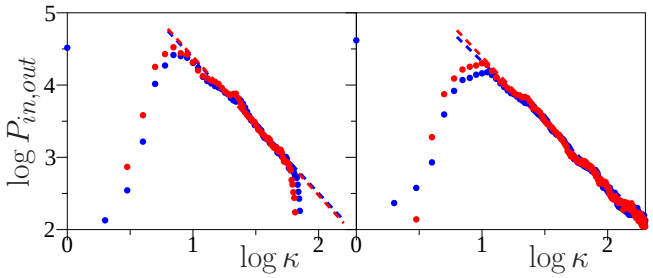

FIG. 7: (Color online) Same as in Fig. 3 for the set $T10$ at $k = 0.22$ (left) and $k = 0.6$ (right) and $N = 3.6 \cdot 10^5$. The fit gives the exponent $\mu = 1.87, 1.92$ for *ingoing* (blue), *outgoing* (red) links at $k = 0.22$ (left) and $\mu = 1.70, 1.83$ for *ingoing* (blue), *outgoing* (red) links at $k = 0.6$ (right).

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.

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

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 $\eta = 1$ [6], the actual value determined numerically is $h = 0.0851$. For $\eta = 0.99$ we also have $\gamma_c = -T \ln \eta = 0.1005$; the theoretical value of the fractal exponent for data in Fig. 5 inset is $\nu = 1 - \gamma_c/(Th) = 1 - 0.118 = 0.882$ while the numerical fit gives $\nu = 0.85$.

For the set $T20$ ($k = 0.3$, $\eta = 0.97$) we have $h = 0.29k^{2/3} = 0.1299$, the actual numerical value is $h = 0.1081$, $\gamma_c = -T \ln \eta = 0.609$; the theoretical fractal exponent for data in Fig. 5 inset is $\nu = 1 - \gamma_c/(Th) = 1 - 0.282 = 0.718$ while the numerical fit gives $\nu = 0.61$.

Figs. 7-10 presented in Appendix give more information for the main part of the paper.

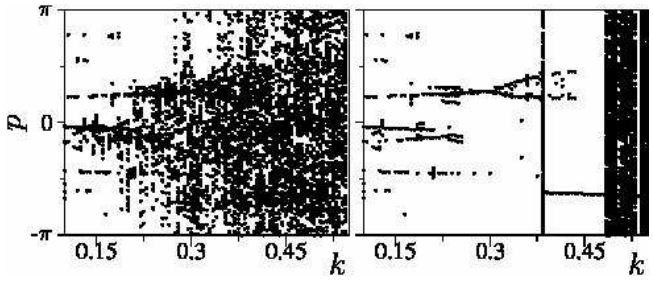

FIG. 8: (Color online) Bifurcation diagram showing values of p vs. map parameter k for the set $T10$. The values of p, obtained from 10 trajectories with initial random positions in the phase space region, are shown for integer moments of time $100 < t/T \leq 110$ (left) and $10^4 < t/T \leq 10^4 + 100$ (right).

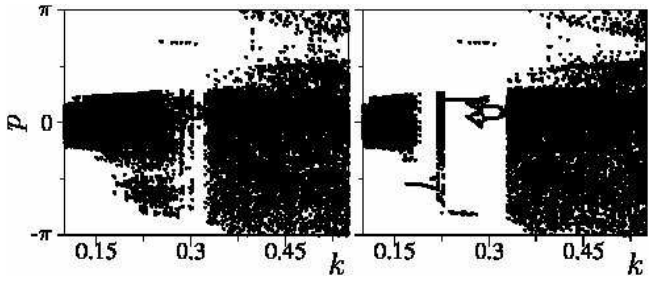

FIG. 9: (Color online) Same as in Fig. 8 for the set $T20$.

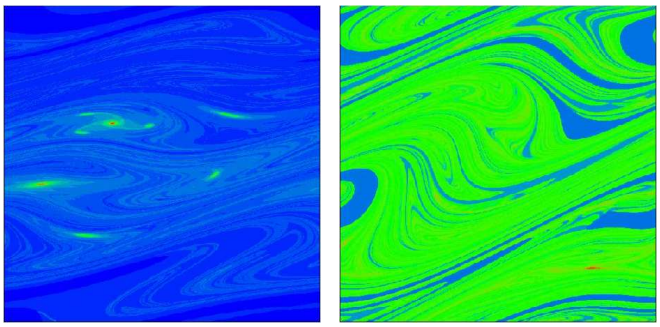

FIG. 10: (Color online) Same as Fig. 1 for the set $T10$ at $\alpha = 0.99$, $N = 3.6 \cdot 10^5$ at $k = 0.22$ (left) and $k = 0.6$ (right); PAR ξ are the same as in the inset of Fig. 6.