Maximal-entropy random walks in complex networks with limited information

Roberta Sinatra,[1,2] Jesús Gómez-Gardeñes,[3,4] Renaud Lambiotte,[5] Vincenzo Nicosia,[1,2] and Vito Latora[1,2]

[1]*Dipartimento di Fisica e Astronomia, Università di Catania and INFN, Via Santa Sofia, 64, I-95123 Catania, Italy*
[2]*Laboratorio sui Sistemi Complessi, Scuola Superiore di Catania, Via San Nullo 5/i, I-95123 Catania, Italy*
[3]*Departamento de Fisica de la Materia Condensada, Universidad de Zaragoza, E-50009 Zaragoza, Spain*
[4]*Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, E-50009 Zaragoza, Spain*
[5]*Institute for Mathematical Sciences, Imperial College London, 53 Prince's Gate, London SW7 2PG, United Kingdom*

Maximization of the entropy rate is an important issue to design diffusion processes aiming at a well-mixed state. We demonstrate that it is possible to construct maximal-entropy random walks with only *local information* on the graph structure. In particular, we show that an almost maximal-entropy random walk is obtained when the step probabilities are proportional to a power of the degree of the target node, with an exponent α that depends on the degree-degree correlations and is equal to 1 in uncorrelated graphs.

In the last decade increasing attention has been devoted to the study of random walks on complex topologies [1,2]. Various features of random walks on networks, such as passage times [3–5] and spectral properties [6,7], have been investigated, and random walks have also been used to detect communities [8,9], to evaluate centrality of nodes [3,10,11], and to obtain coarse-grained graphs [6]. Another quantity recently considered is the entropy rate, a measure to characterize the mixing properties of a stochastic process [12]. In particular, attention has been focused on designing random walks with *maximal entropy rate* on a given graph [13–17], that is, choosing the transition probabilities of the random walk in such a way that the random walkers are maximally dispersing in the graph, exploring every possible walk with equal probability. Practical examples where the maximization of entropy rate is important are diffusion processes which aim at well mixing, such as spreading information about a node's state (its healthy or infected condition, its availability or congestion, etc.) [17], mixing in metapopulations models [18], or global synchronization of moving agents by local entrainment [19].

In principle, the optimization of entropy rate could require the definition of transition probabilities relying on the history of the walker's positions. However, it has been proven that allowing a long-term memory of the past is not needed in order to construct maximal-entropy random walks, since it turns out that there always exists an optimal set of transition probabilities that is Markovian [13–16]. Namely, the maximum entropy rate is obtained with a Markov random walk in which the probability to step from node i to node j is equal to $\frac{a_{ij}u_j}{\lambda u_i}$, where λ is the largest eigenvalue of the adjacency matrix $A = \{a_{ij}\}$ of the graph and **u** is the associated eigenvector [16]. The corresponding value of the maximum entropy rate is equal to $\ln \lambda$. This random walk process has the interesting property to be biased, in the sense that a walker follows a link (i,j) with a probability proportional to the importance of its end j, as measured by its eigenvector centrality u_j [20]. However, the main problem with a real implementation is that, at each time step, the walker needs to have a global knowledge of the network: It needs to know the adjacency matrix of the entire graph. Such global information is very

often unavailable. A walker at a node i usually has only local information, in the sense that it knows the first neighbors of node i and possibly some of their topological properties, such as their degree [17]. In this paper, we prove that almost maximal-entropy random walks can indeed be obtained with a limited and local knowledge of the network. We show how to construct them by solely using the degrees of first and second neighbors of the current node.

Let us consider a Markov random walker moving from node to node on a connected, undirected, and unweighted graph with N nodes and K links. At each time step, the walker moves from the current node to one of its neighbors. If we indicate as $\pi(j|i)$ the probability of jumping from i to j [with the normalization condition $\sum_j \pi(j|i) = 1 \ \forall i$], then $\pi(j|i) \neq 0$ if and only if (iff) $a_{ij} = 1$, that is, if and only if j belongs to the neighborhood of i, $\mathcal{N}_i$. We assume that the walker has the freedom to select the first neighbors of i with different probabilities, so that not all nodes in $\mathcal{N}_i$ are equally selected and some of them are preferred to the others. The simplest case to consider is that of a *regular graph*, that is, a graph in which all nodes have the same degree. The graph can be a random regular graph, or a regular lattice. In such a case, since all the first neighbors of a node are equivalent, the best choice is to set $\pi(j|i) = a_{ij}/\sum_j a_{ij}$, that is, to select one of the nodes in $\mathcal{N}_i$ at random with uniform probability. Things are different in graphs where the nodes have different degrees, especially in graphs with highly heterogeneous degree distributions. As we show below, in these cases, maximally random walks on a graph can be also obtained with only a limited and local knowledge of the topology of the graph.

The optimal random walk on a given graph can be rigorously determined on mathematical grounds by considering the entropy rate h of the stochastic processes associated to different random walks [12]. A trajectory of t steps generated by a random walk starting at a fixed node i is described by the sequence of occupied nodes $i,i_1,i_2,\ldots,i_t$, where $i_1,\ldots,i_t$ are all indices that can take integer values between 1 and N. This means that the walker first moves from i to node i_1, then it jumps to node i_2, and so on. In practice, there is a maximum of $M(t)$ different allowed sequences of length t, corresponding to all possible walks of length t (and starting at node i) on

the graph under study. Depending on the rules of the random walk, not all possible sequences will appear, while some of them will occur with a probability higher than the others. If we denote as joint probability $p(i,i_1,i_2,\ldots,i_t)$ the probability that the sequence $i,i_1,i_2,\ldots,i_t$ is generated by a given random walk, then the entropy rate of the random walk, h, is defined as

$$h = \lim_{t\to\infty} \frac{S_t}{t}, \qquad (1)$$

where S_t is the Shannon entropy of the set of trajectories of length t starting at node i: $S_t = -\sum_{i_1,i_2,\ldots,i_t} p(i,i_1,\ldots,i_t) \ln p(i,i_1,\ldots,i_t)$. The minimum possible value of the entropy rate, $h_{\min} = 0$, is obtained when, for large time t, only one trajectory dominates. On the other hand, the maximum possible value is obtained when, for large time t, all the $M(t)$ allowed trajectories have equal probability to occur, that is, $p(i,i_1,\ldots,i_t) = 1/M(t)$ if $i,i_1,\ldots,i_t$ is a walk on the graph originating in i, and $p(i,i_1,\ldots,i_t) = 0$ otherwise. The maximum value of the entropy is equal to $h_{\max} = \lim_{t\to\infty} \frac{M(t)}{t}$. Now, in the most general case, the probability of having a sequence of t nodes originating at a given node i can be written (for any $t > 1$) in terms of conditional probabilities as

$$p(i,i_1,\ldots,i_t) = p(i_1|i)p(i_2|i,i_1)\ldots p(i_t|i,i_1,\ldots,i_{t-1}).$$

Summing both ends over $i_2,i_3,\ldots,i_t$, and using the normalization conditions $\sum_{i_t} p(i_t|i,i_1,i_2,\ldots,i_{t-1}) = 1$ for $t \geqslant 2$, we get an expression for the conditional probability at the first step as a function of the t times joint probabilities:

$$p(i_1|i) = \sum_{i_2,i_3,\ldots,i_t} p(i,i_1,\ldots,i_t). \qquad (2)$$

This means that, no matter how long the memory in the random walker is, we can always describe it as a Markov random walker, provided that we define the transition matrix of the Markov chain $\pi(i_1|i)$ in terms of the joint probabilities $p(i,i_1,\ldots,i_t)$ as in Eq. (2). In particular, if we want to construct a maximal-entropy random walk, we have to set $p(i,i_1,i_2,\ldots,i_t) = 1/M(t)$ iff $i,i_1,i_2,\ldots,i_t$ is a walk on the graph and $p(i,i_1,i_2,\ldots,i_t) = 0$ otherwise. The number of walks of length t originating in i can be written in terms of the adjacency matrix as $M(t) = \sum_{i_1,i_2,\ldots,i_t} a_{ii_1} a_{i_1 i_2} \ldots a_{i_{t-1} i_t}$. Hence, the joint probability of a trajectory $i,i_1,i_2,\ldots,i_t$ reads

$$p(i,i_1,\ldots,i_t) = \frac{a_{ii_1} a_{i_1 i_2} \ldots a_{i_{t-1} i_t}}{\sum_{i_1,i_2,\ldots,i_t} a_{ii_1} a_{i_1 i_2} \ldots a_{i_{t-1} i_t}}, \qquad (3)$$

and the transition matrix of the Markov random walker with the maximal entropy is finally given by

$$\pi(i_1|i) = \lim_{t\to\infty} \frac{a_{ii_1} \sum_{i_2} a_{i_1 i_2} \ldots \sum_{i_t} a_{i_{t-1} i_t}}{\sum_{i_1} a_{ii_1} \sum_{i_2} a_{i_1 i_2} \ldots \sum_{i_t} a_{i_{t-1} i_t}}. \qquad (4)$$

The value of the entropy rate in Eq. (1) can then be calculated directly from matrix π, as for any ergodic Markov chain, from [12]

$$h = -\sum_{i,j} \pi(j|i) \cdot w^*(i) \ln[\pi(j|i)], \qquad (5)$$

where $w^*(i)$ is the ith component of the stationary distribution. From Eq. (4) it is clear that, in the most general case, in order

for a walker at a node i to select one of its first neighbors to step on, the walker needs to know not only which node is in $\mathcal{N}_i$, but also the neighborhood of first neighbors, the neighborhood of second neighbors, and so on. In practice, the local choice of moving from i to one particular neighbor i_1, depends on the whole adjacency matrix of the graph. However, as we demonstrate below, this global information is not necessary in most of the cases.

Uncorrelated networks. Uncorrelated graphs can be described by the degree sequence of the nodes $\{k(1),k(2),\ldots,k(N)\}$, corresponding to a degree distribution P_k, since the degree of a node does not depend on the degree of its first neighbors. In mathematical terms, this means that the conditional probability $P_{k'|k}$ does not depend on k and can be written in terms of the degree distribution as $P_{k'|k}^{\text{unc}} = k'P_{k'}/\langle k\rangle$, where the right-hand side is the probability to end up in a node of degree k' by choosing an edge at random with uniform probability. Consequently, the average degree of the neighbors of node j, $k_{nn}(j) = 1/k(j)\sum_l a_{jl}k(l)$, does not depend on the degree of j, $k_{nn}(j) = k_{nn}\forall j$, and the last two summations in the numerator and in the denominator of Eq. (4), namely, $\sum_{i_{t-2}} a_{i_{t-3}i_{t-2}} \sum_{i_{t-1}} a_{i_{t-2}i_{t-1}}k(i_{t-1}) = \sum_{i_{t-2}} a_{i_{t-3}i_{t-2}}k(i_{t-2})k_{nn}(i_{t-2})$ can be written as $k_{nn}\sum_{i_{t-2}} a_{i_{t-3}i_{t-2}}k(i_{t-2})$. The constant k_{nn} at the numerator and at the denominator cancels out so that the same argument can be repeated again and again. Finally, the formula factorizes into

$$\pi^1(i_1|i) = \frac{a_{ii_1}k(i_1)}{\sum_{i_1} a_{ii_1}k(i_1)}, \qquad (6)$$

where, by the symbol π^1 we mean the first order approximation to the transition matrix π in Eq. (4). This formula tells us that the best diffusion process on a uncorrelated graph is a random walk whose motion is linearly biased on node degrees. Thus, a walker at a given node, only needs to have information on its first neighbors and their degree. Since the degrees of different nodes are not correlated, a local information of the degree of first neighbors is, in this case, sufficient to construct the diffusion process with maximal entropy. Such information is "locally available" to the walkers, meaning that a walker at node i has complete information on the degree of each node in its neighborhood $\mathcal{N}_i$. Now, it is intuitive that a random walk choosing a node j proportionally to $k(j)$, so that all the trajectories of length 2 starting in i will occur with the same probability, will be more random than a walker selecting uniformly the first neighbors of i.

Formula (6) gives theoretical grounds to the results of Ref. [17], where random walks with power law dependence $\pi(i_1|i) \propto k^\alpha(i_1)$ were explored as a function of α ($\alpha > 0$ indicates a bias toward high-k neighbors, while $\alpha < 0$ means preferring low-k nodes), and it was numerically found that $\alpha = 1$ is the best value if the graph is uncorrelated. Of course, if all nodes have the same degree, as in a regular graph, the transition matrix reduces to that of an unbiased walker:

$$\pi^0(i_1|i) = \frac{a_{ii_1}}{\sum_{i_1} a_{ii_1}}. \qquad (7)$$

This is the lowest possible approximation for π in Eq. (4): In the case of no available information, each neighbor has the same probability to be selected. The values of h obtained numerically with transition matrices π^0 and π^1 in different

TABLE I. The entropies of random walks with no information, $h(\pi^0)$, and with local information, respectively, on nearest, $h(\pi^1)$, and next-nearest neighbors, $h(\pi^2)$, are compared to the maximal possible entropy $h_{\max} = h(\pi) = \ln \lambda$ on different graph models with $N = 500$ and average degree $\langle k \rangle = 6$, on $N = 40 \times 40$ regular square lattices with defects (see Ref. [21]), and on various real networks.

	$\frac{h(\pi^0)}{h(\pi)}$	$\frac{h(\pi^1)}{h(\pi)}$	$\frac{h(\pi^2)}{h(\pi)}$	$h_{\max} = h(\pi)$
Regular lattice	1.000	1.000	1.000	1.79
Random regular graph	1.000	1.000	1.000	1.79
ER random graph	0.954	0.993	0.998	1.98
Uncorrelated scale-free $\gamma = 1.5$	0.886	0.992	0.996	2.36
BA model	0.825	0.976	0.996	2.52
Assortative scale-free $\gamma = 1.5$	0.876	0.991	0.999	2.44
Disassortative scale-free $\gamma = 1.5$	0.937	0.990	0.997	2.18
Regular lattice (1% defects)	0.996	0.997	0.998	1.38
Regular lattice (10% defects)	0.967	0.978	0.981	1.34
Regular lattice (20% defects)	0.931	0.955	0.963	1.29
Internet autonomous system [22]	0.744	0.900	0.980	4.10
U.S. Airports [18]	0.879	0.990	0.997	3.88
E-mail [23]	0.881	0.983	0.997	3.03
SCN (cond-mat) [24]	0.694	0.867	0.946	3.17
SCN (astro-ph) [24]	0.784	0.941	0.973	4.41
PGP [25]	0.597	0.920	0.976	3.75

models of uncorrelated networks are reported in Table I. In agreement with our predictions, in regular lattices and in random regular graphs, $h(\pi^0)$ is equal to the maximal possible entropy $h_{\max} = \ln \lambda$. In Erdős-Rényi (ER) random graphs not all nodes have the same degree, so that a random walk linearly biased on degree has an entropy $h(\pi^1)$ that is much closer to the maximum than $h(\pi^0)$. This effect is even more evident in scale-free graphs, that is, in graphs with a very heterogeneous degree distribution.

Networks with degree-degree correlations. Graphs with degree-degree correlations are described in terms of their degree distribution P_k and of a nontrivial $P_{k'|k}$. This is because the probability that a link from a node of degree k arrives at a node of degree k' does not simply factorize in terms of the degree distribution. In such graphs the average degree of the first neighbors of a node j, $k_{nn}(j)$, does depend on $k(j)$. Therefore, in analogy with Eq. (6) we can define a second order approximation of Eq. (4),

$$\pi^2(i_1|i) = \frac{a_{ii_1} \sum_{i_2} a_{i_1 i_2} k(i_2)}{\sum_{i_1} a_{ii_1} \sum_{i_2} a_{i_1 i_2} k(i_2)}$$
$$= \frac{a_{ii_1} k(i_1) k_{nn}(i_1)}{\sum_{i_1} a_{ii_1} k(i_1) k_{nn}(i_1)}, \quad (8)$$

describing a Markov walker that, at each time step, selects a first neighbor, i_1, of the current node, with a probability proportional to the sum of the degrees of the first neighbors of i_1. This is equivalent to make equiprobale all the walks of length 3 originating in i. In conclusion, to construct high-entropy random walks on correlated graphs, a walker at a given node needs to know the degree of first and second neighbors of the current node, which is still local information.

In Table I we report $h(\pi^2)$ for various models and for real networks. In models of uncorrelated graphs $h(\pi^2)$ is not very different from $h(\pi^1)$, while in models of correlated graphs, in lattices with defects and in most of the networks

from the real world $h(\pi^2)$ is a much better approximation of $h(\pi)$ than $h(\pi^1)$. In most real-world networks, degree-degree correlations are such that the average degree of the first neighbors of a node exhibits a clear power-law dependence on degree: $k_{nn}(j) \sim [k(j)]^{-\nu}$, with $\nu > 0$ ($\nu < 0$) for disassortative (assortative) networks [2]. For instance, as shown in the inset of Fig. 1, $\nu \simeq 0.4$ for the Internet at the autonomous systems (AS) level [22]. Plugging this dependence into Eq. (8), we get an approximate form for the maximal-entropy random walk in a correlated random graph in terms of degree-biased random walks:

$$\pi^2(i_1|i) \simeq \frac{a_{ii_1} [k(i_1)]^{1-\nu}}{\sum_{i_1} a_{ii_1} [k(i_1)]^{1-\nu}}. \quad (9)$$
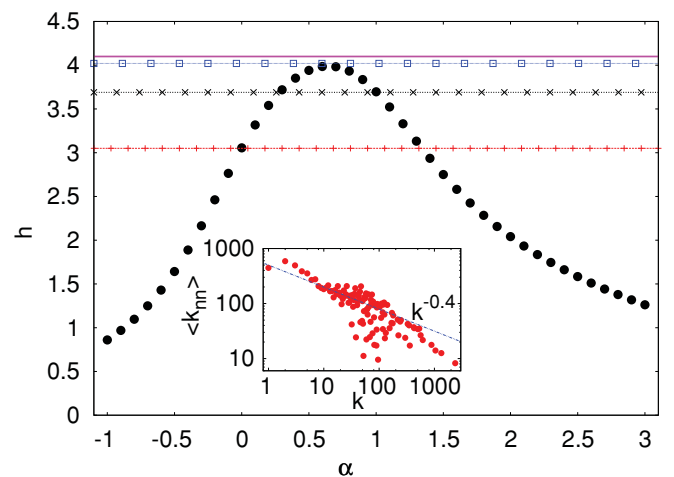

FIG. 1. (Color online) Entropy rate of power-law biased random walks as a function of the degree exponent α for Internet AS. Horizontal lines correspond to, from bottom to top, $h(\pi^0)$, $h(\pi^1)$, $h(\pi^2)$, and $h_{\max} = h(\pi)$. (Inset) Average degree k_{nn} of the first neighbors of nodes of degree k, as a function of k, with fit $k^{-0.4}$.

In practice, on a correlated network, an approximation for the maximal-entropy random walk can be obtained by considering a random walk whose motion is biased as a power of the target node degree, with an exponent $\alpha = 1 - \nu$. Hence, the optimal bias α_{opt} is larger (smaller) than 1 for assortative (disassortative) networks, meaning that we have to prefer a superlinear (sublinear) bias on the node degree. As an example, in Fig. 1 we report the entropy rate of a biased random walk as a function of the exponent α on a disassortative real-world network. We found $\alpha_{\mathrm{opt}} = 0.6$ for Internet AS, which is perfectly in agreement with the value $\nu = 0.4$ in the inset, through the relation $\alpha_{\mathrm{opt}} = 1 - \nu$. We have also checked that this relation holds for the other real networks in Table I.

Networks with higher-order degree correlations. Similar arguments can be repeated for networks with higher-order correlations. This procedure generates a class of biased random walks defined by the transition matrices π^0, π^1, π^2, etc., incorporating more and more information about the system structure. In the supplemental material [21] we studied how this sequence of transition matrices converges to π in different networks. In the limit case in which a graph has correlations at all orders, we have to rely on the full transition matrix of Eq. (4), which can be also expressed by means of the eigenvalues and eigenvectors of the adjacency matrix of the graph. In fact, the numerator and the denominator of (4) can be rewritten in terms of powers of the adjacency matrix, respectively, as $a_{ii_1} \sum_{i_t} (A^{t-1})_{i_1 i_t} = a_{ii_1} (A^{t-1} \cdot \mathbf{1})_{i_1}$ and $\sum_{i_t} (A^t)_{ii_t} = (A^t \cdot \mathbf{1})_i$, where $(A^t)_{ij}$ indicate the entry i, j of matrix A^t, and $\mathbf{1}$ is a vector of ones. By making use of the power method for $t \to \infty$, we finally get

$$\pi(i_1|i) = \frac{a_{ii_1} u_{i_1}}{\lambda u_i} = \frac{a_{ii_1} u_{i_1}}{\sum_j a_{ij} u_j}, \qquad (10)$$

where λ and vector $\mathbf{u}$ are, respectively, the largest eigenvalue and its associated eigenvector of the adjacency matrix [26]. Equation (10) represents a Markov walk whose transition probability is linearly biased by the components of eigenvector $\mathbf{u}$, also known as the eigenvector centrality of the node [20], and it is indeed the same transition matrix proposed in [16] as the process with the maximum possible entropy rate $h_{\max} = \ln \lambda$ [13–16].

In order to perform a maximal-entropy Markov walk on a graph, at each time step, a walker needs a global knowledge of the whole network and has to compute $\mathbf{u}$, which has $O(K)$ computational complexity. However, global information is in practice always unavailable in real systems. As we have shown in this paper, this global knowledge is not necessary since in many real-world networks long-range interactions are weak and can be neglected. It is therefore possible to construct almost maximal-entropy random walks with only local information on the graph structure. This can be done with $O(\langle k \rangle)$ complexity, a dramatic improvement which opens up to practical applications in social, biological, and technological systems.

[1] P. Blanchard and D. Volchenkov, *Random Walks and Diffusions on Graphs and Databases* (Springer, Berlin, 2011).

[2] S. Boccaletti *et al.*, Phys. Rep. **424**, 175 (2006).

[3] J. D. Noh and H. Rieger, Phys. Rev. Lett. **92**, 118701 (2004).

[4] S. Condamin *et al.*, Nature (London) **450**, 77 (2007).

[5] A. Fronczak and P. Fronczak, Phys. Rev. E **80**, 016107 (2009).

[6] D. Gfeller and P. De Los Rios, Phys. Rev. Lett. **99**, 038701 (2007).

[7] V. Zlatic, A. Gabrielli, and G. Caldarelli, Phys. Rev. E **82**, 066109 (2010).

[8] J.-C. Delvenne, S. N. Yaliraki, and M. Barahona, Proc. Natl. Acad. Sci. USA **107**, 12755 (2010).

[9] M. Chavez, M. Valencia, V. Navarro, V. Latora, and J. Martinerie, Phys. Rev. Lett. **104**, 118701 (2010).

[10] L. da Fontoura Costa *et al.*, Appl. Phys. Lett. **91**, 054107 (2007).

[11] S. Lee, S.-H. Yook, and Y. Kim, Eur. Phys. J. B **68**, 277 (2009).

[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).

[13] W. Parry, Trans. Am. Math. Soc. **112**, 55 (1964).

[14] L. Demetrius and T. Manke, Physica A **346**, 682 (2005).

[15] J. C. Delvenne and A.S. Libert, e-print arXiv:0710.3972v2.

[16] Z. Burda, J. Duda, J. M. Luck, and B. Waclaw, Phys. Rev. Lett. **102**, 160602 (2009).

[17] J. Gómez-Gardeñes and V. Latora, Phys. Rev. E **78**, 065102(R) (2008).

[18] V. Colizza, R. Pastor-Satorras, and A. Vespignani, Nat. Phys. **3**, 276 (2007).

[19] N. Fujiwara, J. Kurths, and A. Díaz-Guilera, Phys. Rev. E **83**, 025101(R) (2011).

[20] P. Bonacich, J. Math. Soc. **2**, 113 (1972).

[21] See supplemental material at [http://link.aps.org/supplemental/10.1103/PhysRevE.83.030103] for details about the quality of the approximations shown in Table I.

[22] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).

[23] R. Guimera, L. Danon, A. Diaz-Guilera, F. Giralt, and A. Arenas, Phys. Rev. E **68**, 065103(R) (2003).

[24] M. E. J. Newman, Phys. Rev. E **64**, 016131 (2001).

[25] M. Boguña, R. Pastor-Satorras, A. Diaz-Guilera, and A. Arenas, Phys. Rev. E **70**, 056122 (2004).

[26] For uncorrelated graphs, $u_j \sim k_j$ [27], and Eq. (10) reduces to Eq. (6).

[27] J. G. Restrepo, E. Ott, and B. R. Hunt, Phys. Rev. E **76**, 056119 (2007).