Collective Behavior of a Small-World Recurrent Neural System with Scale-Free Distribution

Zhidong Deng, *Member, IEEE,* and Yi Zhang

Abstract

This paper proposes a scale-free highly-clustered echo state network (SHESN). We designed the SHESN to include a naturally evolving state reservoir according to incremental growth rules that account for the following features: (1) short characteristic path length, (2) high clustering coefficient, (3) scale-free distribution, and (4) hierarchical and distributed architecture. This new state reservoir contains a large number of internal neurons that are sparsely interconnected in the form of domains. Each domain comprises one backbone neuron and a number of local neurons around this backbone. Such a natural and efficient recurrent neural system essentially interpolates between the completely regular Elman network and the completely random echo state network (ESN) proposed by H. Jaeger *et al*. We investigated the collective characteristics of the proposed complex network model. We also successfully applied it to challenging problems such as the Mackey-Glass dynamic system and the laser time series prediction. Compared to the ESN, our experimental results show that the SHESN model has a significantly enhanced echo state property and better performance in approximating highly complex nonlinear dynamics. In a word, this large-scale dynamic complex network reflects some natural characteristics of biological neural systems in many aspects such as power law, small-world property, and hierarchical architecture. It should have strong computing power, fast signal propagation speed, and coherent synchronization.

Index Terms

Echo state network, local preferential attachments, recurrent neural networks, scale-free, small-world, time series prediction.

Manuscript received December 10, 2004. This work was supported in part by the National Science Foundation of China under Grant No. 60321002 and the Teaching and Research Award Program for Outstanding Young Teachers in Higher Education Institutions of MOE (TRAPOYT), China.

Z. D. Deng is with the Department of Computer Science, Tsinghua University, Beijing 100084, China. He is also with the State Key Laboratory of Intelligent Technology and Systems, Tsinghua University, Beijing 100084, China.

Y. Zhang is with the Department of Computer Science, Tsinghua University, Beijing 100084, China.

Collective Behavior of a Small-World Recurrent Neural System with Scale-Free Distribution

I. Introduction

In a large number of real-world complex networks, such as the nervous system of *C. Elegans*, cellular and metabolic networks, food webs, the World-Wide Web, the BBS, the Internet backbone, power grids, citation networks, and many social networks, small-world phenomena and scale-free properties have been discovered in the past few years [1], [2], [12], [19], [28], [30], [31]. The small-world network first introduced by Watts and Strogatz [31] is defined as a highly clustered network with a short characteristic path length, and was conceptually originated from Milgram's work in the late 1960s, or the so-called "six degrees of separation" principles [24]. Differing from homogeneous or exponential networks, the scale-free distribution of network connectivity, which is defined as ratio of edges to possible edges, has a power law form [1]. This implies that the majority of the nodes in a scale-free network have a sparse connection; only the minority have a dense connection. The unveiling of a small-world effect and a scale-free property, which are viewed as universal laws governing different kinds of complex networks, including biological networks [2] and brain functional networks [10], [31], has directly led to dramatic advances in exploring large-scale complex networks. On the other hand, biological neural systems typically have a massive number of neurons, recurrent pathways, sparse random connectivity, and local modification of synaptic weights [17], [18], which are very different from those of most of the existing artificial neural network models such as BP networks [27], Hopfield networks [14], and Vapnik's SVMs (support vector machines) [6]. In general, traditional artificial neural networks, ranging from feedforward and feedback to recurrent structures, use at most several dozen neurons in various practical applications. The number of neurons in hidden layers is often chosen subjectively according to different problems under consideration. They usually have a slow learning speed and suboptimal solution. Recently, several new neural network models that have some characteristics of complex network topology have attracted the attention of many researchers. Associative memory neural networks with either small-world architecture [5], [20] or scale-free topology [9], [15] have been presented, all of which demonstrate better performance in memory capacity and time than randomly connected Hopfield networks with the same connectivity. Similarly, for the synchronization problem of chaotic dynamic system, either small-world effect or scale-free feature can make synchronization more efficient and robust [3], [13], [16], [21], [29]. All their investigations show that the two complexities effects have a dramatic influence upon the collective dynamic behavior of complex system.

Interestingly, H. Jaeger and his co-authors proposed a new artificial recurrent neural network (RNN), called

echo state networks (ESNs), for function approximation, chaotic time series prediction, and modeling of nonlinear dynamic system [17], [18]. The ESN model contains a completely random state reservoir as a hidden layer, which is usually composed of hundreds or thousands of internal neurons. This very promising RNN partially reflects some features of learning mechanisms in biological brains and the resulting learning algorithm is computationally efficient and easy to use. Remarkably, the accuracy of predicting a chaotic time series using the ESN is significantly enhanced by a factor of 2,400 over previous techniques [18]. In order to have such surprising capabilities of approximating nonlinear dynamic system, however, the ESN must satisfy so-called echo state property. According to the definition given by Jaeger [17], the ESN has echo state property only if the current network state is uniquely determined by the history of inputs after running for a long time. This means that for the ESN the spectral radius of reservoir weight matrix must be not greater than 1 [17], [18]. It seemingly makes sense that the state error equation of the ESN's reservoir starting with different initial states will converge when the eigenvalues of reservoir weight matrix are all located within the unit circle in the z-plane. But this sufficient condition may considerably restrain the enhancement of the approximation capabilities of this model. In fact, when the ESN is used to model highly nonlinear systems, the computing power of the ESN is directly correlated with its memory capacity, or more exactly, to its massive short-term memory [18]. As mentioned by [17], the larger the spectral radius, the slower is the decay of the network's response to an impulse input, and the stronger the network's memory capacity. In this case, the ESN can have a more efficient computing power and a better approximation capability. In other words, the spectral radius has a full impact on approximation capabilities of the ESN. But for the ESN model, the spectral radius of the reservoir is not allowed to be greater than 1 so that the echo state property holds. Therefore, it is highly expected to relax this sufficient condition for the spectral radius of the state reservoir, in order to overcome the limitation of approximating strongly nonlinear dynamic systems using the ESN.

Inspired by all the above results, we investigate the possibility of introducing various complexities to the state reservoir of the ESN model so that the resulting model is able to reflect more learning mechanisms of biological brain. In fact, we examined a state reservoir that only has small-world features but no scale-free characteristics. Unfortunately, approximation capabilities of the resulting small-world ESN model are quite poor. Furthermore, introducing scale-free distribution alone also has little improvement in performance. These observations, the demand of further improvement of approximation capabilities, and many neurobiological advances, have, in part, motivated our research work presented here.

We propose an extended ESN model that has a scale-free and small-world state reservoir. The new state reservoir comprises thousands of internal neurons that are sparsely interconnected. In this paper, the emphasis is on our natural-growth rules and analysis of the complexities of networks. The proposed scale-free highly-clustered complex network was successfully applied to the Mackey-Glass dynamic system and the laser time series prediction problem. Our empirical results showed that this new complex network model greatly improves the echo state property by allowing a larger range of viable spectral radius and considerably enhances performance of approximating highly complex dynamic systems.

II. SHESN: SCALE-FREE HIGHLY-CLUSTERED ECHO STATE NETWORK

Unlike either the completely regular Elman networks [11] or completely random ESNs introduced by H. Jaeger *et al.* [17], [18], we propose a *s*cale-free *h*ighly-clustered *e*cho *s*tate *n*etwork (SHESN) that has a naturally evolving state reservoir. The architecture of the SHESN is shown in Fig. 1, and consists of three layers: an input layer, a new state reservoir (or a hidden dynamic layer), and an output layer. The new state reservoir is generated by incremental growth. Embedded in the SHESN's structure, our new state reservoir network has a collection of naturally evolving features, including scale-free or power law distribution for node degree, high clustering coefficient, short characteristic path length, and hierarchical and distributed architecture.

A. Network Architecture of SHESN

The network architecture of the SHESN is shown in Fig. 1. At time step k, the input layer receives an r-dimensional input vector $u(k) = [u_1(k), u_2(k), \cdots, u_r(k)]^T$ and passes it to all the internal neurons in the new state reservoir through an $n \times r$ input weight matrix W^{in}. The i-th neuron in this layer is directly connected to the i-th component $u_i(k)$ of input vector $u(k)$. The output of the i-th input neuron simply equals its corresponding input $u_i(k)$. In the new state reservoir, all n internal neurons are sparsely interconnected using an $n \times n$ reservoir weight matrix W^{res}. The output of each internal neuron is called a state, and all the states are denoted by $x(k) = [x_1(k), x_2(k), \cdots, x_n(k)]^T$. The neurons in the output layer collect all the states $x_i(k)(i = 1, 2, \cdots, n)$ in the preceding new state reservoir, along with all the inputs $u_i(k)(i = 1, 2, \cdots, r)$ in the first layer, through an $m \times (n+r)$ output weight matrix W^{out}, which will be discussed later, and consequently produce an m-dimensional output vector $y(k) = [y_1(k), y_2(k), \cdots, y_m(k)]^T$ of the whole SHESN. Meanwhile, the network output vector $y(k)$ is fed back to all the internal neurons through an $n \times m$ feedback weight matrix W^{fb}. The activation functions in the last two layers are all set to the $tanh(\ )$ function. Just like that of [18], input and feedback weight matrices of W^{in} and W^{fb} are randomly assigned with uniform distribution, and the output weight matrix W^{out} is adjusted using supervised learning. But the reservoir weight matrix W^{res} is produced according to our naturally evolving rules rather than the completely random approaches used in [17], [18].

The forward propagation of SHESN can be summarized as:

$$x(k+1) = \tanh(W^{res}x(k) + W^{in}u(k+1) + W^{fb}y(k) + v(k)) \quad (1)$$

$$y(k) = \tanh(W^{out} \begin{bmatrix} x(k) \\ u(k) \end{bmatrix}), \quad (2)$$

where $v(k)$ is a noise (threshold) that is added to the activation functions of internal neurons.

The framework of the SHESN implements an adaptive dynamic system using structures analogous to other recurrent neural networks, which enables the proposed SHESN at least to gain the advantage of a fast learning speed and capabilities of approximating nonlinear dynamic systems.

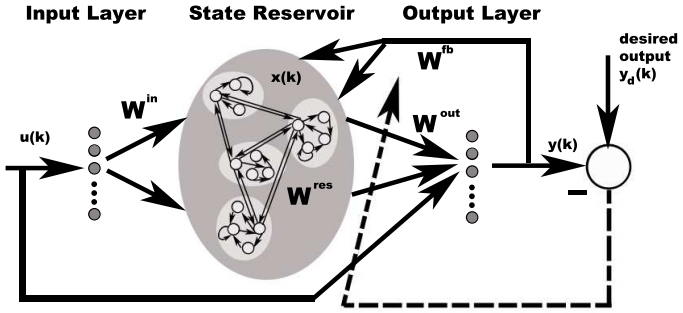


Fig. 1. System architecture of SHESN. The generated state reservoir has a collection of naturally evolving features.

B. Natural Growth Model of the SHESN's Reservoir

Now we concentrate on the natural growth rules of the SHESN's reservoir. In fact, some interesting natural features of biological brains, such as the small-world property and the scale-free feature [10], have not been reflected in the randomly connected ESN reservoir. However, many investigations show that neural network models that have either small-world effect or scale-free distribution demonstrate excellent performance in memory capacity, time, and synchronization [3], [5], [9], [13], [15], [16], [20], [21], [29]. To our knowledge, there has not yet published any neural network that simultaneously contains the two complexities characteristics according to natural growth rules.

Additionally, biological networks and many other complex networks also have a hierarchical and distributed architecture. For instance, neurons in brains are organized into many functional columns such as cerebral cortex. The Internet network generally contains many domains, each of which is composed of a large number of local nodes. To get such an ubiquitous architecture, the basic idea underlying the BRITE model, a parameterized Internet topology generator proposed by A. Medina and his co-authors [23], is employed to design our natural growth rules that produce the SHESN's reservoir. Note that in the BRITE model, both the initial assignments of constructive backbones and the use of the preferential attachment are critical to the incremental growth of scale-free networks. But it is impossible to generate small-world networks using the BRITE.

Inspired by the BRITE, the initial topology of our SHESN's reservoir is designed to contain a set of fully connected backbone neurons, each of which represents a potential domain. We then present a set of natural growth rules in order to generate the SHESN's reservoir that has both small-world and scale-free features, which is expected

to achieve superior approximation capabilities over the ESN.

The procedure of generating a naturally evolving state reservoir roughly involves the following six steps:

a) Initialize an $L \times L$ grid plane of a state reservoir.

b) Put backbone neurons on the grid plane and then generate synaptic connections among the backbone neurons.

c) Add one new local neuron to the grid plane.

d) Produce synaptic connections for the newly added local neuron using local preferential attachment rules.

e) Repeat steps c) and d) for each new local neuron.

f) Generate a reservoir weight matrix W^{res} such that the echo state property holds.

As mentioned above, a network has echo state property only if the network state $x(k)$ is uniquely determined by any left-infinite input sequence $[\cdots, u(k-1), u(k)]$. More precisely, this indicates that for any given left-infinite input sequence, for all state sequences $[\cdots, x(k-1), x(k)]$ and $[\cdots, x'(k-1), x'(k)]$, which are updated using (1), it holds that $x(k) = x'(k)$, or, $e(k) = x(k) - x'(k) = 0$. Hence the echo state property implies that the internal neurons in the state reservoir show systematic variations of the driver signal.

1) Initialization of an $L \times L$ Grid Plane of a State Reservoir: Let the number of internal neurons in a new state reservoir be n, where $n \ll L \times L$. Using a stochastic dynamic growth model like that proposed by [23], the internal neurons with a size of n are incrementally assigned on a grid plane of state reservoir divided into $L \times L$ squares. The $L \times L$ grid plane or the reservoir with a capacity of $L \times L$ is required to be large enough in order to contain all the n internal neurons ($n \ll L \times L$). Note that different internal neurons gradually generated can not be placed at the same location on the grid plane of the reservoir and are not allowed beyond the grid plane.

2) Generation of Backbone Neurons and Synaptic Connections Associated: In the process of generating a new state reservoir, we classified all the internal neurons into two categories: backbone neurons and local neurons. Usually, the number of backbone neurons must be much less than that of local neurons. In our experiment described later, for example, about one percent of internal neurons in the new state reservoir were backbone ones. Specifically, we randomly generated the x- and y-coordinates [x_b, y_b] of each backbone neuron on the grid plane. A fixed set of backbone neurons was then assigned at different locations on the $L \times L$ grid plane, which roughly determined the spatial distribution of all n internal neurons. Furthermore, we defined a domain as the set of internal neurons that comprises one backbone neuron and a number of local neurons around this backbone. A collection of domains constituted our new state reservoir from a higher level perspective.

The spatial distribution of n_b backbone neurons, however, must satisfy two restrictions. One is that different backbone neurons generated are not allowed to be at the same location on the grid plane of the reservoir. The other is that the minimum distance between any two backbone neurons must be greater than a certain threshold such that the resulting domains could be separated from each other. For instance, we set this threshold to be 30 in our experiment. After that, the backbone neurons of n_b were fully connected to each other through synaptic connection weights that were randomly set to be a real value between [-1, 1].

3) Incremental Growth of New Local Neurons: As described above, we initially built a small fully connected backbone network, which consisted of n_b backbone neurons. Other (local) neurons were then generated so as to

form an entire naturally evolving state reservoir. In this process, for each local neuron, we randomly select one of the backbone neurons and put the local neuron into the domain associated with this backbone. Suppose that the location of this backbone neuron is [x_b, y_b]. We then produce the x- and y-coordinates of local neurons around the backbone one [x_b, y_b] on the grid plane according to the bounded Pareto heavy-tailed distribution [7]:

$$P(v) = \frac{ak^a}{1-(k/Q)^a} v^{-a-1}, \quad k \le v \le Q,$$

where a denotes the shape parameter and $k(Q)$ the minimum (maximum) value.

If a local neuron is farther from its own backbone neuron than one of the other backbone neurons, we assume this local neuron is located in another domain dominated by the nearest backbone neuron, which enables the incremental growth of new local neurons to avoid collisions. But this case, fortunately, happens rarely in the procedure presented here due to the fact that the bounded Pareto distribution seldom generates a high value [7].

As a result, it causes a popular and natural effect that in the same domain, or equivalent cortex in a sense, most of the local neurons are spatially near to their backbone, while a few of them are far away from it. Such a spatial distribution of internal neurons bears a resemblance to that which occurs in the human brain network [10].

4) Generation of Synaptic Connections for New Local Neuron Using Local Preferential Attachment Rules: Based on preferential attachment rules [1], any newly added local neurons always prefer to connect to neurons that already have many synaptic connections. More precisely, the probability that a new local neuron is connected to an existing neuron is proportional to the outdegree of the existing neuron. Considering the domain concept introduced here, we present a new strategy, henceforth called a local preferential attachment.

For convenience, we refer to the domain that contains a new local neuron as the current domain. The candidate neighborhood of a new local neuron is defined as the set of neurons to which this new local neuron is allowed to be connected. Specifically, assume that there exists a circle whose center is the location of the new local neuron and the radius is the Euclidean distance from the new local neuron to the backbone neuron in the current domain. Consequently, we choose all those neurons in the current domain that are just within such a circle as the candidate neighborhood of the new local neuron. Apparently, the backbone neuron is always one of the candidate neighbors farthest from the new local neuron in the current domain.

Let n_c represent the number of synaptic connections for a newly added local neuron. The parameter n_c controls the density of connections generated in the current domain. Additionally, let n_1 and n_2 be the number of neurons, respectively, in the current domain and in the candidate neighborhood of a new local neuron ($n_1 \ge n_2$).

Our local preferential attachment rules are given as follows:

(a) If $n_c \ge n_1$, a new local neuron is fully connected to all the existing neurons in the current domain.

(b) If $n_2 \le n_c < n_1$, the candidate neighborhood is redefined as the set of all the neurons in the current domain, instead of just those neurons in the circle specified above. A new local neuron is connected to all the candidate neighbors using the following probability [23]:

$$\frac{d_i}{\sum_{j \in C} d_j},$$

where d_j is the current outdegree of neuron j, and C the candidate neighborhood of the new local neuron.

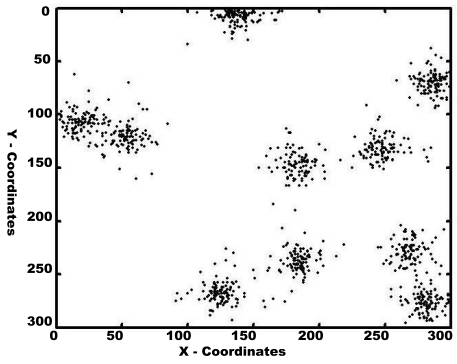


Fig. 2. The spatial distribution of 1,000 neurons on a 300 × 300 grid plane.

(c) If $n_c < n_2$, the probability that is used to attach a new local neuron to a candidate neighbor is the same as the above.

In general, local preferential attachment rules contribute to both the small-world and the scale-free properties of generated networks. The chosen candidate neighbors of a new local neuron, of which the new local neuron itself is usually in the center, should help improve the clustering coefficient of the network.

C. Analysis of the Complexities of the SHESN

Based on the natural growth model presented above, we incrementally generated a state reservoir of the SHESN. The network topology parameters are as follows: the capacity of the reservoir $L \times L$ = 300×300 = 90,000, the number of internal neurons n = 1,000, the number of backbone neurons $n_b = 10$, and the number of connections for each local neuron $n_c = 5$. In addition, we took the parameters of the bounded Pareto heavy-tailed distribution as follows: a = 1, k = 1, and Q = 90,000. The spatial distribution of the naturally evolving reservoir produced is shown in Fig. 2. The spectral radius and the sparse connectivity of such a new reservoir were calculated to be 2.1051 and 0.979%, respectively.

1) Spatially Hierarchical and Distributed Structure: As shown in Fig. 2, the 1,000 internal neurons are incrementally placed on the 300×300 grid plane, resulting in 10 clear clusters or domains. Each domain contains one backbone neuron, and a collection of local neurons around the backbone one. It is readily evident that the network topology of our naturally evolving reservoir has a spatially hierarchical and distributed structure at different levels. It has several unique features, as follows:

(a) The new dynamic state reservoir comprises several domains. The domains are regarded as top level macro-neurons in the reservoir network hierarchy. They are fully connected to each other through backbone neurons. The number of inter-domain synaptic connections among backbone neurons is much smaller than the number of entire connections in the new reservoir due to the fact that there is a small percentage of backbone neurons.

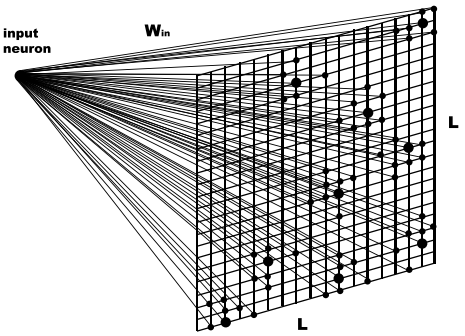


Fig. 3. The input weight matrix W^{in} describing the relationship between one input neuron and all the internal neurons.

(b) In each domain or at a low level, local neurons are only connected to the neurons located in the same domain. In most cases, the number of intra-domain connections is much greater than that of inter-domain ones.

(c) According to (a) and (b), dynamic behaviors for each domain are relatively independent.

(d) All the n internal neurons in the resulting reservoir are spatially scattered on an $L \times L$ grid plane ($L \times L \gg n$), as shown in Fig. 3. Initially, each neuron in the input layer is connected to all the $L \times L$ nodes in the grid plane. When the n internal neurons in the grid plane are generated according to the natural growth rules, the input weight matrix between each input neuron and the n internal neurons can be randomly produced with uniform distribution. Meanwhile, all the input weights between each input neuron and the grid nodes that are not internal neurons are reset to zero. Hence the input weight matrix reflects the spatial distribution of internal neurons.

2) Small-World Phenomenon: Average Characteristic Path Length and Clustering Coefficient: The average characteristic path length $L(p)$ and the clustering coefficient $C(p)$ [31] are used to characterize the small-world phenomenon of complex network topology. Our naturally evolving reservoir network was composed of 1,000 internal neurons. Correspondingly, it had a 1,000×1,000 reservoir weight matrix W^{res}, with a sparse connectivity of 0.979%. Hence it was undoubtedly a large and sparse complex network.

As a global property, the average characteristic path length $L(p)$ reflects the effective size of a complex network. It is defined as the mean distance between two internal neurons, averaged over all pairs of internal neurons. The clustering coefficient $C(p)$ is a local property measuring the probability that two neurons neighboring an internal neuron are also neighbors of each other [31]. More precisely, suppose that the internal neuron v in the reservoir has k_v neighbor neurons that are connected to the neuron v. Let n_a (n_t) represent the actual (total possible) number of connections among the neighbor neurons of v. The clustering coefficient C_i of the neuron v is defined as $C_i = n_a/n_t = 2n_a/(k_v(k_v - 1))$ [31]. Hence the clustering coefficient $C(p)$ of the whole reservoir network is the average of C_i over all the internal neurons.

For the SHESN's reservoir presented above, the average characteristic path length and the clustering coefficient were computed as follows: $L(p)$ = 3.7692 and $C(p)$ = 0.2303, respectively. For comparison, we investigated a

TABLE I

ANALYSIS OF SMALL-WORLD AND SCALE-FREE FEATURES FOR TEN DOMAINS

Domain	1	2	3	4	5	6	7	8	9	10
Size	103	108	88	91	104	86	99	114	85	122
Char. Path Length	2.0041	2.0043	1.9432	1.9761	1.9678	1.9886	1.9486	2.0723	1.9837	2.0399
Clustering Coef.	0.4082	0.4691	0.4653	0.4569	0.3957	0.4392	0.4558	0.3701	0.4167	0.3955
Correlation Coef.	0.9875	0.9843	0.9909	0.9850	0.9901	0.9891	0.9843	0.9847	0.9852	0.9821

completely random reservoir network with the same size of 1,000 neurons and a sparse connectivity of about 1%. For this ESN reservoir network, we calculated the average characteristic path length and the clustering coefficient, respectively, *i.e.*, L_{random} = 3.2668 and C_{random} = 0.0112. These calculations showed that $L(p)$ was almost as small as L_{random} and $C(p)$ was much greater than C_{random}. In other words, the new reservoir of the SHESN had short average characteristic path lengths and high clustering coefficients as well. Therefore our new reservoir is a small-world complex network.

Moreover, Table I gives the average characteristic path length and the clustering coefficient for each of the 10 domains. Compared to their counterparts in the completely random network, we showed that each domain is also a small-world subnetwork.

3) Scale-Free Feature: It is well known that power laws are free of any characteristic scale. Networks that have power law degree distributions are called scale-free networks. In recent years, empirical studies have revealed that Internet topologies exhibit power laws in the form of $y = x^{-\alpha}$ [12], [23], [26]. Usually, the power law exponent α is employed to characterize some universal properties of network topology. In order to find the exponent α, we first plotted the relationship between two variables x and y in a log-log scale and the slope of the resulting linear plot was then viewed as the power law exponent α. Furthermore, we utilized Pearson's correlation coefficients for the log-log plot so as to judge whether power laws really exist. In fact, the closer the absolute value of the correlation coefficient is to one, the more closely the data obey a power law [23]. For a good linear fit on a log-log plot, the correlation coefficient should often be greater than 0.95 and the p-value should be less than 0.05 as well. In general, if the p-value is small, then the correlation is significant.

Let us consider the following two power law or scale-free distributions [12]: outdegree of neurons *vs*. rank and the number of neurons *vs*. outdegree. Both of them were observed in our naturally evolving reservoir. Note that the rank of a neuron is defined as the order of a neuron in all the internal neurons that are sorted in descending order according to their outdegrees.

Fig. 4 shows the relationship between the outdegree of internal neurons and the rank of neurons on a log-log plot. Using the *corrcoef* function in MATLAB, the correlation coefficient was calculated to be 0.988 with the p-value of 0. The rank exponent R, or the slope of the fitting linear plot, was calculated as 0.59, which is in accordance

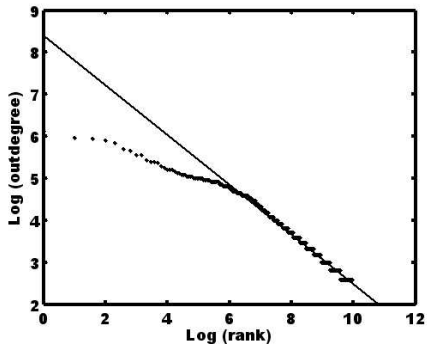


Fig. 4. Log-log plot of outdegree of neurons *vs*. rank. The correlation coefficient is 0.988 with a p-value of 0.

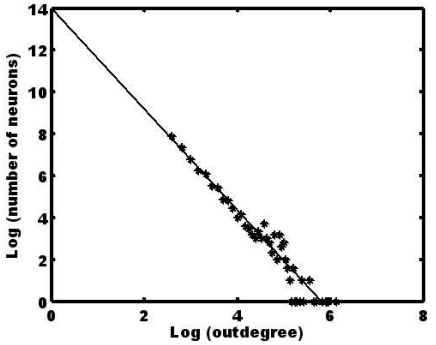


Fig. 5. Log-log plot of the number of neurons vs. outdegree. The correlation coefficient is 0.979 with a p-value of 0.

with the results for the Internet topology in [12].

Furthermore, we also found the correlation coefficient for each domain, as listed in Table I. The results illustrate that every domain, *i.e.*, low-level subnetworks, also exhibits the power law characteristic.

Similarly, as shown in Fig. 5, we obtained the relationship between the number of internal neurons and the outdegree. Note that we eliminated the neurons that have outdegree outliers before linearly fitting the log-log plot. The correlation coefficient was calculated as 0.979 and the p-value 0. In this case, we computed the outdegree exponent of reservoir O = 2.41. Actually, the outdegree exponents of 2.0 – 3.0 has frequently been discovered in most natural complex networks such as biological networks [1], [30]. This illustrates that our SHESN expresses some biological characteristics at least in terms of power law distribution [10].

D. Training the SHESN Using Supervised Learning

As mentioned above, the connection weight matrix of a reservoir, $i.e.$, W^{res}, must be carefully chosen in order to ensure that the echo state property is retained. Meanwhile, the input weight matrix W^{in} and the feedback weight W^{fb} could be arbitrarily assigned within the range of possibilities. The output weight matrix W^{out}, however, must be adjusted using supervised learning.

In other words, after constructing the recurrent neural network architecture that satisfies the echo state property, we must compute the output weight matrix W^{out} such that the training sample dataset is approximated by the network.

For instance, the training dataset of n_t length is defined as follows:

$$\{u(1); y_d(1)\}, \{u(2); y_d(2)\}, \cdots, \{u(n_t); y_d(n_t)\}$$

where $u(k)$ denotes the input vector at time k and $y_d(k)$ the desired output.

After discarding the first n_0 steps or going through the transient time [18], we must find the output weight matrix W^{out} such that the training mean-square error (*MSE*):

$$MSE = 1/(n_t - n_0) \sum_{k=n_0+1}^{n_t} \left(d(k) - W^{out} \begin{bmatrix} x(k) \\ u(k) \end{bmatrix} \right)^2$$

is minimized. Note that $d(k) = (\tanh)^{-1}(y_d(k))$, $x(k) = \begin{bmatrix} x_1(k) & x_2(k) & \cdots & x_n(k) \end{bmatrix}^T$, and n indicates the number of echo state variables.

Apparently, this is a classical linear regression problem. It is easy to directly solve it using a generalized inverse matrix approach. Therefore the $1\times(n+1)$ dimensional output matrix W^{out} is derived as follows:

$$W^{out} = (M^{-1}D)^T,$$

where T represents the transpose. The $(n_t - n_0) \times (n + 1)$ dimensional matrix is given by

$$M = \begin{bmatrix} x_1(n_0+1) & \cdots & x_n(n_0+1) & u(n_0+1) \\ x_1(n_0+2) & \cdots & x_n(n_0+2) & u(n_0+2) \\ \cdots & \cdots & \cdots & \cdots \\ x_1(n_t) & \cdots & x_n(n_t) & u(n_t) \end{bmatrix}$$

and

$$D = \begin{bmatrix} d(n_0+1) & d(n_0+2) & \cdots & d(n_t) \end{bmatrix}^T.$$

In our experiment, calculations of a generalized inverse matrix of M were done by using the *pinv* pseudoinverse function in MATLAB.

III. Enhanced Echo State Property and Dynamic Approximation Capability

A. Dataset Preparation

1) The Mackey-Glass System: The Mackey-Glass dynamic system, with large time delay τ, is a well-known testbed for nonlinear chaotic system prediction. As an example, we first applied the SHESN to this challenging problem in order to validate performance of our new network model. The differential equation of the Mackey-Glass system is as follows [22]:

$$\frac{dx}{dt} = \frac{0.2x(t-\tau)}{1 + x(t-\tau)^{10}} - 0.1x(t),$$

TABLE II

SEVERAL TRAINING AND TEST DATASETS FOR THE MG SYSTEM

Sampling Dataset	Time Delay τ	Transformation
Dataset 1	17	tanh(x–1)
Dataset 2	30	(0.3tanh(x–1))+0.2
Dataset 3 – 17	17 – 31	(0.3tanh(x–1))+0.2

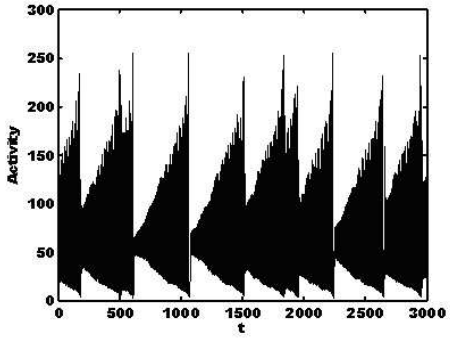


Fig. 6. The first 3,000 data points in the laser time series.

where x represents a state and τ a time delay. Apparently, the larger the time delay τ, the more severe the nonlinearity of the MG system. Specifically, a chaotic time series occurs in the MG system when $\tau \geq 17$.

In order to prepare training and test datasets, we used the *dde23* function in MATLAB to solve the delay differential equation. In particular, we set absolute accuracy to $1e$-16 instead of a default value of the *dde23* function. As a consequence, seventeen sampling datasets were constructed using the above equation solver as the time delay τ increased from 17 to 31, as listed in Table II. Note that transformations were done to set all the generated samples to be within available output ranges. For the sake of comparison, dataset 2 was the same as that used in the supporting online material of [18].

2) Laser Time Series Prediction: The laser time series problem, as a real-world chaotic time series, has extensively been used to test a variety of prediction approaches [4], [8], [18], [25], [32]. As a second example, we used the same dataset as the one in [18]. Fig. 6 shows the first 3,000 data points of the laser time series. Note that there are several breakdown events that are very difficult to predict.

Similarly, we constructed two training and test datasets as follows: (1) Dataset 18: The first 2,200 data points of the laser time series were used to form the training dataset. The test dataset comprised the segment between 2,200 to 2,400. (2) Dataset 19: We used the first 2,500 data points as an alternative training dataset. The data points between 2,500 to 2,700 were selected as the test dataset in this case. It is readily observed from Fig. 6 that there exists a breakdown event in each of the two test datasets.

B. Testing Criterion

Here we give the testing criterion for the SHESN (and the ESN) in our experiments. In the test phase, we evaluated the accuracy of the SHESN (and the ESN) either at a specific observation point or for all the data points on 100 independent runs.

Specifically, for the MG system, the 84 samples following the training dataset of n_t length, *i.e.*, $\{u(n_t+1); y_d(n_t+1)\}$, $\{u(n_t+2); y_d(n_t+2)\}$, $\cdots$, $\{u(n_t+84); y_d(n_t+84)\}$, were chosen as a test dataset [18]. Similarly, we took $k = n_t + 84$ as an observation point and completed 100 independent trials. Note that for each trial we used a newly generated SHESN. Then we could compute a normalized root-mean-square error (NRMSE) as our test error:

$$NRMSE_{84} = \left(\sum_{l=1}^{100} \left(y_d^l(n_t+84) - y^{\,l}(n_t+84) \right)^2 / \sum_{l=1}^{100} \left(y_d^l(n_t+84) \right)^2 \right)^{1/2}, \tag{3}$$

where $y^{\,l}(k)$ was the actual output of the SHESN at the l-th trial.

For the laser time series prediction problem, instead of selecting one observation point, we used all of the 200 data points in a test dataset, e.g, from $n_t + 1$ to $n_t + 200$, to compute the NRMSE for those data points on 100 independent runs as follows:

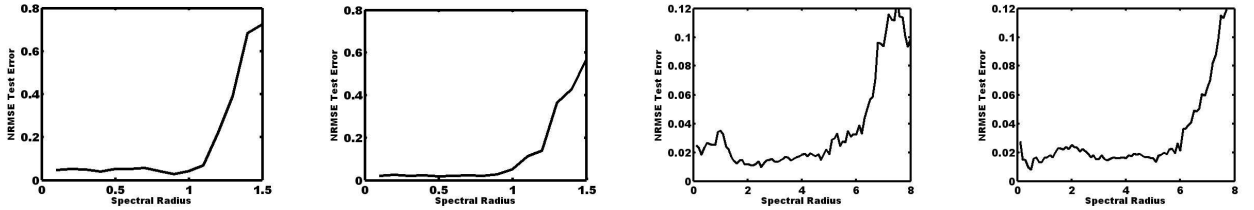
$$NRMSE = \left(\sum_{l=1}^{100} \sum_{m=n_t+1}^{n_t+200} \left(y_d^l(m) - y^{\,l}(m) \right)^2 / \sum_{l=1}^{100} \sum_{m=n_t+1}^{n_t+200} \left(y_d^l(m) \right)^2 \right)^{1/2}. \tag{4}$$

C. Enhanced Echo State Property

In the completely random ESN approach proposed by Jaeger *et al*. [18], the state reservoir is sparsely and randomly connected and all the internal and output neurons have sigmoid activation functions such as *tanh*. The output of internal neurons in the reservoir is referred to as a state. Using the nonlinear state equation described above, the state sequence of the reservoir $X = \{\cdots, x(k-1), x(k)\}$ is correspondingly produced as we sequentially provide a set of left-infinite time series $U = \{\cdots, u(k-1), u(k)\}$ as an input of the network. If the reservoir has an asymptotically stable state, the state of reservoir $x(k)$ is called the echo state. In general, the ESN has an echo state property only if a state of reservoir $x(k)$ is uniquely determined by any left-infinite input sequence $[\cdots, u(k-1), u(k)]$ [17]. In other words, the echo state property means that the ESN has an echo state in an admissible state set X with input from U.

Let $|\lambda_{\max}|\ (W^{res})$ denote the spectral radius of the connection weight matrix of the state reservoir, where $\lambda_{\max}$ is an eigenvalue of W^{res} with the largest magnitude. For such a randomly connected ESN, Jaeger pointed out that the sparse connection matrix of the ESN reservoir W^{res} must satisfy $|\lambda_{\max}|\ (W^{res}) < 1$ so that the state of the reservoir $x(k)$ can act as the echoing function of the driving signal [18]. In fact, it is quite consistent with our empirical studies that the ESN can not work properly when its state reservoir has a spectral radius beyond 1.

As one of the most important contributions of this paper, the SHESN model we propose in this paper indicates that the sufficient condition for the ESN $|\lambda_{\max}|\ (W^{res}) < 1$ is quite conservative. The echo state property of the SHESN can be significantly enhanced by allowing a much wider range of spectral radius. Four experiments on the MG system and the laser time series prediction were done in order to examine echo state properties of the ESN



(a) Testing ESN on dataset 1 (b) Testing ESN on dataset 2 (c) Testing SHESN on dataset 1 (d) Testing SHESN on dataset 2

Fig. 7. The $NRMSE_{84}$ test error *vs*. spectral radius on MG system.

and our SHESN. In the empirical studies below, we used dataset 1, dataset 2, dataset 18, and dataset 19 as the training and testing datasets for four different experiments, respectively.

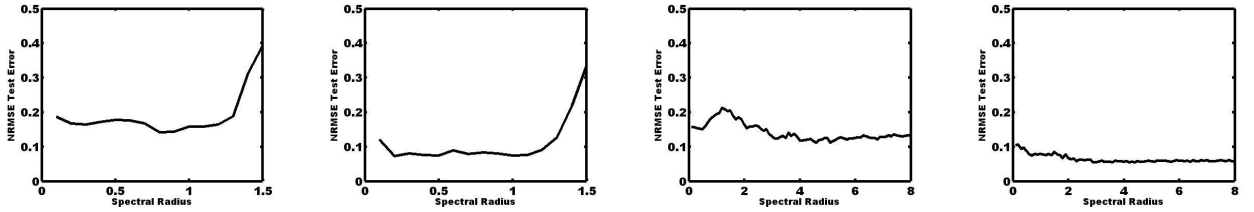
(1) Experiment for the ESN: As stated above, the spectral radius of the connection weight matrix of the ESN reservoir $|\lambda_{\max}|$ (W^{res}) must be less than 1 so as to satisfy the echo state property. In other words, the ESN will have an asymptotically unstable null state if $|\lambda_{\max}| > 1$ [17].

Here we carried out four experiments, based on dataset 1, dataset 2, dataset 18, and dataset 19, respectively. For datasets 1 and 2, we constructed an ESN reservoir of 500 internal neurons with sparse connectivity of 1%. The noise $v(k)$ added here was randomly produced with uniform distribution over [-0.0008, 0.0008]. The input and feedback weight matrices W^{in} and W^{fb} were randomly set with uniform distribution over [-1, 1]. The output weight matrix W^{out} was derived using the supervised learning algorithm described above. We then examined the $NRMSE_{84}$ test errors, which could be calculated by (3), with the increase of the spectral radius starting with 0.1 at a stepsize of 0.1. The experimental results are given in Fig. 7(a) and Fig. 7(b), respectively.

For datasets 18 and 19, we used the reservoir with 500 internal neurons for the ESN, the input and feedback weight matrices W^{in} and W^{fb} were randomly set with uniform distribution over [-1, 1] and [-0.4, 0.4], respectively, according to private communications with H. Jaeger [17], [18]. The sparse connectivity was assigned to 2% in this case. Then we investigates the $NRMSE$ test errors, which could be calculated by (4), with the increase of the spectral radius starting with 0.1 at a stepsize of 0.1. Empirical results are shown in Fig. 8(a) and Fig. 8(b).

It is clear from Fig. 7(a), Fig. 7(b), Fig. 8(a), and Fig. 8(b) that all of the four ESNs became unstable when the spectral radius of the reservoirs was greater than 1. This is completely consistent with the results of [18]. In particular, we observed from Fig. 7(b) that when the spectral radius was 0.8, the $NRMSE_{84}$ test error was calculated as 0.0185, which was consistent with that obtained in [18]. Note that for experiments on dataset1 and 2, we used the same training and test datasets and similar network parameters, compared to the experiments done in [18].

(2) Experiment for the SHESN: Similar to the four different experiments done for the ESN, we chose the reservoir topology parameters of the SHESN. For dataset 1 and 2, the capacity of reservoir $L \times L$ = 200×200 = 40,000, the number of internal neurons n = 500, the number of backbone neurons $n_b = 5$, and the number of connections for a new local neuron $n_c = 5$. The noise $v(k)$ was randomly generated with uniform distribution



(a) Testing ESN on dataset 18 (b) Testing ESN on dataset 19 (c) Testing SHESN on dataset 18 (d) Testing SHESN on dataset 19

Fig. 8. The $NRMSE$ test error *vs.* spectral radius on laser time series.

over [-0.0008, 0.0008]. The input and feedback weight matrices W^{in} and W^{fb} were randomly set with uniform distribution over [-1, 1]. The output weight matrix W^{out} was derived using the supervised learning algorithm. The experimental results are shown in Fig. 7(c) and Fig. 7(d), respectively.

For datasets 18 and 19, the maximum capacity of the reservoir with 500 internal neurons was set as $L \times L$ = 200×200 = 40,000. We selected the number of backbone neurons $n_b = 5$ and the number of connections for a new local neuron $n_c = 1$. The input and feedback weight matrices W^{in} and W^{fb} were randomly set with uniform distribution over [-1, 1] and [-0.4, 0.4], respectively. We eventually achieved the experimental results for both datasets 18 and 19, as shown in Fig. 8(c) and Fig. 8(d), respectively.

The results showed that the SHESN had a much wider range of spectral radius. Even if the spectral radius $|\lambda_{\max}|$ (W^{res}) was significantly greater than 1, the state of SHESN reservoir could still be considered as the echoing functions of the input signal. This considerably improves the echo state property, compared to those results obtained for ESNs (Fig. 7(a), Fig. 7(b), Fig. 8(a), and Fig. 8(b)). As shown in Fig. 7(c), Fig. 7(d), Fig. 8(c), and Fig. 8(d), our observation indicated that the upper boundary of the spectral radius of our new reservoir increased even up to 6.0.

D. Approximating Nonlinear Dynamics

1) The Mackey-Glass System: We first applied the SHESN to model the Mackey-Glass dynamic system with a large time delay. The goal was to examine the SHESN's capability to approximate nonlinear dynamics. In order to compare our SHESN to the ESN, we employed datasets 3 – 17 listed in Table II, in which the time delay varied from 17 to 31 correspondingly. For both the ESN and the SHESN, we took the same network topologies and parameters as those done for the two experiments using dataset 2, respectively. The sole difference was that we designed a different spectral radius for naturally evolving reservoirs. In this experiment, the spectral radiuses of the SHESNs for the MG system using datasets 3 – 17 were all set to 2.7, while the spectral radiuses of the randomly connected ESNs were all set to the Jaeger's original value of 0.8 [17], [18]. The experimental results are shown in Fig. 9.

Note that the reason for choosing the spectral radius of 0.8 for the ESN in our comparative study is that this is the best value for the MG system in a whole range of time delay 17 – 31. More importantly, it is in accord with

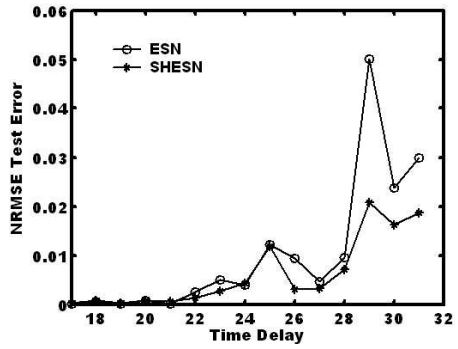


Fig. 9. The NRMSE test error *vs.* the time delay τ in the MG system.

that used in [17], [18].

For the Mackey-Glass dynamic system, the larger the time delay τ, the more severe the system nonlinearity. Specifically, a chaotic time series occurs in the MG system when $\tau \geq 17$. Hence it is very hard to approximate the MG dynamic system using either ESN/SHESN or any other models with the increase of τ. This, undoubtedly, is a real challenge. In Fig. 9, the SHESN error becomes significantly smaller than the ESN when $\tau \geq 26$. In this case, the MG system is of significantly high nonlinearity and needs to have particularly strong capabilities of short-term memory. Our SHESN does achieve a better result. Note that it is unfair to compare the results of both the SHESN and the ESN obtained at two different time delays, because there have quite different difficulties led by different levels of nonlinearity.

From Fig. 9, both the SHESN and the ESN have approximately the same performance when in the range of $\tau = 17 - 25$. But the SHESN has stronger capabilities of approximating nonlinear dynamics than the ESN, as τ is increased up to 26 and beyond, which implies that system nonlinearities become more severe and this problem is very tough to tackle. For example, the superior performance of the SHESN over the ESN is quite apparent as $\tau = 29$. The reason is that the greater the spectral radius, the stronger the computational capabilities of the dynamic reservoir. This causes the new reservoir of the SHESN to be capable of approximating more complex nonlinearities using stable echo states. Further explanations will be offered in the discussion section.

In view of the randomness in both the ESN and SHESN, a measure of variance for performance should be tractable. Fig. 9 shows the NRMSE test error for 100 independent runs. The standard deviations for both the ESN and SHESN are listed in Table III.

2) Laser Time Series Prediction: As shown in Fig. 8(a), Fig. 8(b), Fig. 8(c), and Fig. 8(d), we tested the prediction performance of the ESN and the SHESN on datasets 18 and 19, respectively. Here we compared the prediction accuracy of the ESN and the SHESN on a specific spectral radius. For the ESN, the spectral radius of 0.9 is the best value from Fig. 8(a) and Fig. 8(b). For the SHESN, we chose spectral radius of 4.0, which is much larger than the upper bound of spectral radius of the ESN. The results are listed in Table IV. Note that the NRMSE test errors were calculated according to (4) on 100 independent runs.

TABLE III

STANDARD DEVIATION OF PREDICTION FOR THE MG SYSTEM

Time Delay	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
ESN ($\times 10^{-4}$)	0.032	0.159	0.053	0.181	0.068	0.276	0.398	0.117	0.794	0.233	0.271	0.624	2.551	0.968	2.755
SHESN($\times 10^{-4}$)	0.037	0.149	0.031	0.267	0.048	0.384	0.364	0.171	1.769	0.171	0.337	0.847	0.689	0.889	2.191

TABLE IV

THE NRMSE TEST ERRORS FOR THE LASER TIME SERIES

	dataset 18	dataset 19
ESN	0.1422	0.0804
SHESN	0.1148	0.0558

IV. DISCUSSION

For the ESN, it has been shown [17], [18] that the echo state property does not hold if the sparse reservoir matrix W^{res} satisfies $|\lambda_{\max}|\ (W^{res}) > 1$. We have demonstrated in our experiments that for the SHESN with a naturally evolving reservoir, the echo state property can be greatly enhanced by allowing a larger range of viable spectral radius. This, undoubtedly, raises a very interesting question of why it happens. What is the underlying reason for these amazing phenomena?

To explain the essential difference between the new state reservoir of the SHESN and the randomly generated reservoir of the ESN, we examine the eigenvalue spectrum of state reservoirs, which is known to dominate the dynamics of the reservoir networks. We first set up the random reservoir of the ESN and the naturally evolving reservoir of the SHESN, respectively, both of which had 500 internal neurons and a spectral radius of 1.5. The

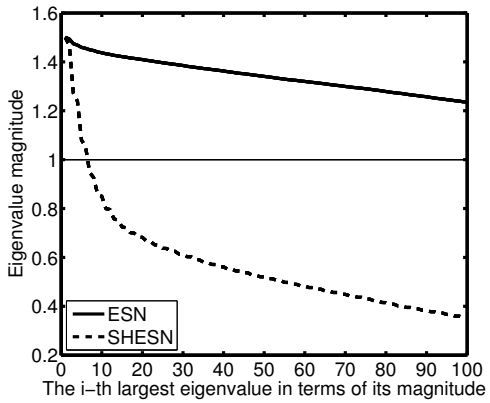


Fig. 10. Spectrum for the largest 100 eigenvalues in terms of their magnitudes as the spectral radius is 1.5.

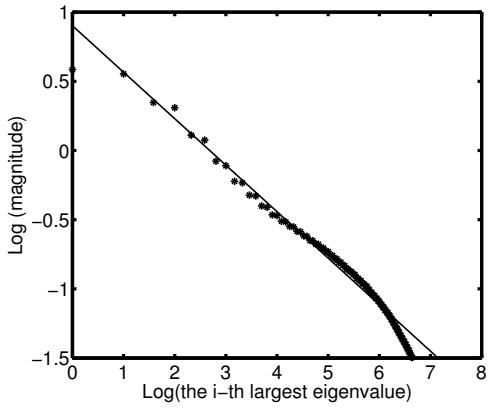


Fig. 11. Log-log plot of the eigenvalue magnitude *vs*. the i-th largest eigenvalue.

connectivity of the ESN's reservoir was set to be 1%. For the SHESN's reservoir, we used the same parameters as those in the laser time series prediction described above. We then investigated the eigenvalue spectrum, namely, the distribution of the eigenvalues in the z-plane for both the reservoirs. We carried out 100 independent runs, and calculated the average of the magnitude of the i-th largest eigenvalue over the 100 independent runs ($i = 1, 2, \cdots, 100$), respectively, as shown in Fig. 10. It is readily observed from Fig. 10 that the magnitudes of the largest eigenvalues (*i.e.*, $i = 1$) for the two reservoirs are all equal to 1.5. For the ESN reservoir, the curve declines very smoothly, and the magnitudes of all the 100 largest eigenvalues are around 1.5. Specifically, among all the 500 eigenvalues, there are 207 eigenvalues that have magnitudes of greater than 1. Meanwhile, the eigenvalue magnitudes of the SHESN's reservoir drop sharply. Only 6 eigenvalues, *i.e.*, 1.2 percent of the 500 eigenvalues, have magnitudes beyond 1. More interestingly, for the SHESN's reservoir, the distribution of 100 largest eigenvalue magnitudes obeys the power-law, where the correlation coefficient is 0.989 and the p-value 0, as shown in Fig. 11. This indicates that the magnitudes of most of the eigenvalues reduce exponentially and is very close to the origin of the z-plane, although the spectral radius of the SHESN's reservoir is greater than 1.

We also investigated whether the performance of the SHESN degrades as random disturbance increases. We tried a wider range of noise levels and obtained very encouraging experimental results. We conducted our experiments with the SHESN based on dataset 19, where the parameters were chosen to be the same as those in the Section III-C. When the spectral radius is set to be 8.0 and the noise disturbance $v(k)$ ranges from $[-10^{-7}, 10^{-7}]$ to $[-10^{-1}, 10^{-1}]$, the approximating error roughly remains unchanged. The system had indeed a very strong robustness. Like the same reason that scale-free networks are robust to random failures [1], our SHESN architecture is more robust to random noise.

Actually, the power law distribution of the eigenvalue magnitudes guarantees that even when the SHESN reservoir has a larger spectral radius, only a few of eigenvalues are beyond the unit circle, whereas most of them have small magnitudes that fall sharply. Within the unite circle, a large number of eigenvalues that are very close to the origin of the z-plane will act as dominant eigenvalues and determine the global convergence of system, even in the case of

larger random noise. Therefore, in the high-dimensional phase space of internal neurons in the SHESN reservoir, a very few large eigenvalues, *i.e.*, non-dominant eigenvalues, span only a much lower-dimensional subspace. In this case, two trajectories may diverge locally and very transiently due to their different initial states and the random noises, they will not reach global divergence.

This research work may raise some deep issues that are not resolved yet. For example, the SHESN is probably quite susceptible to targeted attacks on some specific neurons, particularly on the backbone neurons and the neurons with high degrees. Furthermore, the SHESN architecture may be less robust to noise on specific (hub) neurons. All these open problems, including strict mathematical proofs of the SHESN's enhanced echo state property, are really very interesting topics and worthy of being explored in the future.

V. Conclusion

In this paper we proposed a scale-free highly-clustered echo state network (SHESN), which is an extended ESN model that contains a small-world and scale-free state reservoir. In order to establish the new state reservoir, we presented incremental growth rules that took into consideration several natural characteristics: (1) short characteristic path length, (2) high clustering coefficient, (3) scale-free distribution for network degree, and (4) hierarchical and distributed architecture. We investigated the collective behavior of the SHESN and then applied the SHESN to prediction problems of the MG system and the laser time series. The experimental results obtained showed that compared to the Jaeger ESN with a randomly connected state reservoir, our new SHESN network significantly improves the echo state property and is capable of accurately approximating highly complex nonlinear dynamic systems. Such a natural and efficient recurrent neural system, which could consist of thousands of neurons or even more, is very likely to represent some neurobiological features in many aspects such as scale-free distribution and small-world property. Future work will focus on synchronization problems of such a small-world and scale-free dynamic recurrent neural network. Research along this line and practical applications to other complex dynamic problems are in progress.

Acknowledgment

The authors would like to thank anonymous referees for their useful and very valuable comments on the manuscript.

References

[1] A.-L. Barabasi and R. Albert, *Science*, vol. 286, p. 509, 1999.

[2] A.-L. Barabasi and Z. N. Oltvai, *Nature Rev.*, vol. 5, p. 101, 2004.

[3] M. Barahona and L. M. Pecora, *Phys. Rev. Lett.*, vol. 89, p. 054101, 2002.

[4] G. A. Barreto and A. F. R. Araujo, *IEEE Trans. Neural Netw.*, vol. 15, p. 1244, 2004.

[5] J. W. Bohland and A. A. Minai, *Neurocomputing*, vol. 38, p. 489, 2001.

[6] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines*. London, UK: Cambridge University Press, 2000.

[7] M. Crovella, M. Harchol-Balter, and C. Murta, in *Proceedings of ACM Sigmetrics '98 Conference on Measurement and Modeling of Computer Systems Poster Session*, Madison, WI, 1998.

[8] J. C. P. A. A.-B. D. Erdogmus, O. Fontenla-Romero and E. Castillo, *IEEE Trans. Neural Netw.*, vol. 16, p. 325, 2005.

[9] L. d. F. C. D. Stauffer, A. Aharony and J. Adler, *Eur. Phys. J. B*, vol. 32, p. 395, 2003.

[10] V. M. Eguiluz, D. R. Chialvo, G. Cecchi, M. Baliki, and A. Apkarian, 2005.

[11] J. Elman, *Cognitive Science*, vol. 14, p. 179, 1990.

[12] M. Faloutsos, P. Faloutsos, and C. Faloutsos, in *ACM SIGCOMM*, Cambridge, MA, 1999.

[13] M. Y. C. H. Hong and B. J. Kim, *Phys. Rev. E*, vol. 65, p. 026139, 2002.

[14] J. J. Hopfield and D. W. Tank, *Biological Cybernetics*, vol. 52, p. 141, 1985.

[15] J. M. J. J. Torres, M. A. Munoz and P. L. Garrido, *Phys. Rev. Lett.*, p. 018102, 2005.

[16] G. C. J. Lu, X. Yu and D. Cheng, *IEEE Trans. Cir. and Sys., Part 1*, vol. 51, p. 787, 2004.

[17] H. Jaeger, GMD-Report 148, German National Research Institute for Computer Science, Tech. Rep., 2001. [Online]. Available: ftp://borneo.gmd.de/pub/indy/publications_herbert/EchoStatesTechRep.pdf

[18] H. Jaeger and H. Haas, *Science*, vol. 304, p. 78, 2004.

[19] K. Klemm and V. M. Eguiluz, *Phys. Rev. E*, vol. 65, p. 036123, 2002.

[20] G. A. L. G. Morelli and M. N. Kuperman, *Eur. Phys. J. B*, vol. 38, p. 495, 2004.

[21] X. Li and G. Chen, *IEEE Trans. Cir. and Sys., Part 1*, vol. 50, p. 1381, 2003.

[22] M. C. Mackey and L. Glass, *Science*, vol. 197, p. 287, 1997.

[23] A. Medina, I. Matta, and J. Byers, *ACM SIGCOMM Computer Communication Review*, vol. 30, p. 18, 2000.

[24] S. Milgram, *Psychology Today*, vol. 2, p. 60, 1967.

[25] M. C. P. Tino and L. Benuskova, *IEEE Trans. Neural Netw.*, vol. 15, p. 6, 2004.

[26] Z. L. Ren, Z. D. Deng, D. X. Shuai, and J. Gu, *Computer Physics Communications*, vol. 136, p. 225, 2001.

[27] D. E. Rumelhart and J. L. McClelland, *Parallel distributed processing: explorations in the microstructure of cognition*. Cambridge, MA: MIT Press, 1986.

[28] S. H. Strogatz, *Nature*, vol. 410, p. 268, 2001.

[29] X. F. Wang and G. Chen, *IEEE Trans. Cir. and Sys., Part 1*, vol. 49, p. 54, 2002.

[30] ——, *IEEE Circuits and Systems Magazine*, vol. 3, p. 6, 2003.

[31] D. J. Watts and S. H. Strogatz, *Nature*, vol. 393, p. 440, 1998.

[32] A. S. Weigend and N. A. Gershenfeld, *Time Series Prediction: Forecasting the Future and Understanding the Past.* Reading, MA: Addison-Wesley, 1994.