Echo in a Small-World Reservoir: Time-Series Prediction using an Economical Recurrent Neural Network

Yuji Kawai, Tatsuya Tokuno, Jihoon Park, and Minoru Asada
Graduate School of Engineering, Osaka University
2-1 Yamada-oka, Suita, Osaka, 565-0871 Japan
Email: {kawai, tatsuya.tokuno, jihoon.park, asada}@ams.eng.osaka-u.ac.jp

***Abstract*—A small-world topology has been found in the cortical neural connectivity. However, the role of the topology in neural information processing has yet not been well understood. In this article, we investigate the performance of an echo state network (ESN) within a small-world topology in an economical or cost-effective environment, i.e., reduced number of input/output reservoir nodes. The ESN, a type of recurrent neural network, has a reservoir network where nodes are connected to each other with fixed weights. We introduce the small-world topology into the reservoir network. The ESN learns about the connected weights from the reservoir nodes to an output layer. In order to leverage the potential of the small-world topology, we limit the number of the reservoir nodes that receive external input (i.e., input nodes) or omit their signals to the output layer (i.e., output nodes). In addition, we segregate the input nodes from the output nodes, thereby necessitating the propagation of the input signals to the output nodes through the small-world reservoir. In our experiment, the ESNs learned to predict the next input of chaotic time-series. The small-world ESN exhibited high performance even when the number of input and output nodes was reduced, whereas the performance of the standard random or fully connected ESNs declined with reduced number of nodes.**

I. Introduction

While brain cortical networks are sparse, complex, and economical [1], [2], their information processing is excellent. The nodes constituting the networks are not fully connected with each other and construct sparse complex networks. One of the most famous topology of such networks is the "small-world" topology [3]. Many studies have been conducted to investigate the small-world topology in the anatomical and functional networks of human brains (e.g., [4], [5]), brains of monkeys and cats [6], and nervous systems of *C. elegans* [3] (refer to [7], [8] for reviews). However, the role of the topology in neural information processing has not been well understood.

A small-world network lies between regular and random networks. Watts and Strogatz [3] proposed a method to constitute the network. At first, they constructed a regular network where nodes were connected only with their neighbors (see Fig. 1 (a)). Next, these connections were rewired with a probability p to a randomly selected node. The network for $p = 0$ is still regular, while $p = 1$ results in a random network (Fig. 1 (c)). When p lies between 0.01 and 0.1, it

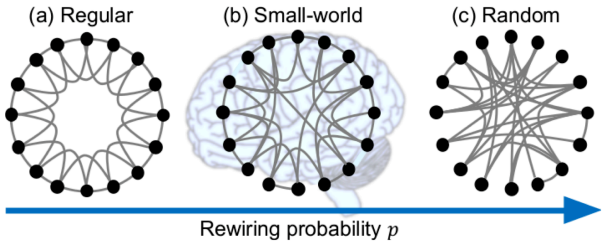


Fig. 1. The small-world model proposed by Watts and Strogatz [3]. The connections of the regular network (a) are rewired with probability p. The network with $p = 0.1$ has a small-world topology (b), and the network with $p = 1$ turns random (c).

is called a small-world network and can spread information with a minimum number of long-range short cuts, resulting in low wiring-costs (Fig. 1 (b)). This type of network is characterized by two factors: the shortest path length, L, and the clustering coefficient, C. L is defined as the average number of connections in the shortest path between two nodes. C is defined as the density of closed triangles, or triplets, consisting of three connected nodes. A regular network is characterized by a large C as well as a large L, while in a random network both L and C are small. In contrast, the small-world network has a small L and a large C.

One approach to understand the potential of the small-world topology is by evaluating the performance of artificial neural networks with respect to the topology. Many researchers have reported that the small-world topology can improve their learning performance [9]–[13]. Simard et al. [9] installed the topology in a multi-layered feedforward network with back-propagation. They rewired some feedforward connections of the regular (standard) feedforward network and demonstrated that the small-world topology improved its learning performance. In addition, the small-world topology can provide a better associative memory for a Hopfield network than the regular or random networks [12], [13]. We would like to investigate the effects of the small-world topology from a more dynamical perspective.

Reservoir computing is a kind of a recurrent neural network [14], [15]. Randomly connected nodes constitute the reservoir,

receive external inputs, and create complex dynamics. All, or some, reservoir nodes emit their state signals to an output layer (readout) through their output connection weights. The greatest feature of reservoir computing is to learn about the output weights in a supervised manner, while other weights between the reservoir nodes are fixed. An echo state network (ESN) is a type of reservoir computing in which the readout is linear [14]. This type of architecture is known as a cortico-striatal model where the reservoir represents the cortex, and the learning for the readout is considered as a function of the striatum [16]. Therefore, this seems to be a reasonable model to investigate the effects of the small-world topology on brain dynamics.

Some studies have already introduced the small-world topology to the reservoir of the ESN and evaluated the performance of the network [17], [18]. However, they concluded that the topology did not significantly improve its learning performance. In this respect, we have assumed that an excessively rich connectivity obscured the advantages of the small-world topology. Their ESNs, as standard ESNs, possessed a direct connection from the input layer to the output layer, and not through the reservoir, and a feedback connection from the output to the reservoir. In addition, all reservoir nodes received signals from the input layer and sent their state to the output layer. Such non-economical configurations might boost their learning performance to a maximum and hide the potential of the topology. It is a fact that in the real brain, not all neurons receive sensory input or emit output, e.g., motor commands. Although it has been known that an ESN with both small-world and scale-free topologies exhibits better learning performance than the standard ESN [19], we focus on the only small-world topology.

We examine the performance of the ESN with the small-world topology in an economical environment to demonstrate the merits of the topology. We limit the number of reservoir input nodes that receive signals from the input layer as well as the output nodes that send signals to the output layer. In addition, we segregate the reservoir input nodes from the output nodes to ensure that the input signals pass through the small-world network. We abolish the direct connections between the input and output layers as well as the feedback connection. In this way, our ESN has fewer connections and, therefore, its learning cost is expected to be lower. We examine how the small-world topology overcomes these economical constraints and realizes an efficient learning.

II. Small-world echo state network

A. Architecture of an echo state network

Fig. 2 shows the proposed ESN with the small-world topology. Some reservoir nodes receive K-dimensional input $\mathbf{u}(t) = (u_1(t), \cdots, u_K(t))^\top$, weighed by a matrix $\mathbf{W}_{\rm in}$ at time t. States of the N reservoir nodes, $\mathbf{x}(t) = (x_1(t), \cdots, x_N(t))^\top$, are updated according to

$$\mathbf{x}(t+1) = f(\mathbf{W}_{\rm in}\mathbf{u}(t+1) + \mathbf{W}\mathbf{x}(t)), \tag{1}$$

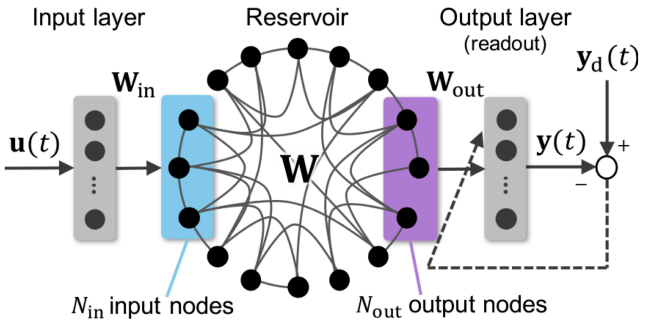


Fig. 2. The echo state network with a small-world topology. The reservoir weight matrix $\mathbf{W}$ represents a ring shaped small-world topology. The part of the reservoir nodes that is enclosed in the blue box receives external input, and the other part that is enclosed in the purple box sends its states to the output layer. The output connection weights in $\mathbf{W}_{\rm out}$ are trained in a supervised manner.

where $\mathbf{W}$ is an $N \times N$ reservoir weight matrix and is constant over the learning phase. We employ the hyperbolic tangent as the function f. Of importance here, is that the only $N_{\rm in}$ ($< N$) reservoir nodes (input nodes) receive $\mathbf{u}(t)$, that is, $\mathbf{W}_{\rm in}$ is sparse. These limited input nodes are enclosed within the blue box in Fig. 2. Similarly, only $N_{\rm out}$ ($< N$) reservoir nodes (output nodes) pass their states $\mathbf{x}_{\rm out}(t)$ into the output layer, shown enclosed within the purple box in Fig. 2. This constraint on the I/O nodes helps clarify the potential of the small-world topology. The M-dimensional output $\mathbf{y}(t) = (y_1(t), \cdots, y_M(t))^\top$ at time t is given as

$$\mathbf{y}(t) = f_{\rm out}(\mathbf{W}_{\rm out}\mathbf{x}_{\rm out}(t)), \tag{2}$$

where $\mathbf{W}_{\rm out}$ is the $M \times N_{\rm out}$ output weight matrix. This matrix is trained with the desired output $\mathbf{y}_{\rm d}(t)$. Since the output function $f_{\rm out}$ is linear, the training is attained using a simple linear regression. After the reservoir is driven T times, we define two matrices: (1) a state collection matrix, $\mathbf{S}$, of dimensions $T \times N_{\rm out}$, where $\mathbf{S} = (\mathbf{x}_{\rm out}(1), \cdots, \mathbf{x}_{\rm out}(T))$ and $\mathbf{x}_{\rm out}(t)$ indicates a state vector of the output nodes, and (2) a desired output collection matrix, $\mathbf{D}$, of dimensions $T \times M$, where $\mathbf{D} = (\mathbf{y}_{\rm d}(1), \cdots, \mathbf{y}_{\rm d}(T))$. The output matrix is computed as

$$\mathbf{W}_{\rm out} = ((\mathbf{S}^\top\mathbf{S})^{-1}\mathbf{S}^\top\mathbf{D})^\top. \tag{3}$$

This model has fewer connections and, thus, is more economical than the previous ones [17]–[19]. We expect that the small-world topology works advantageously in such economical configurations. The differences between this model and the previous models are as follows:

- Limitation on the number of the I/O nodes (i.e., $N_{\rm in}$, $N_{\rm out} < N$). In the previous models, all nodes were capable of receiving input and emitting output (i.e., $N_{\rm in} = N_{\rm out} = N$).
- Segregation between input and output reservoir nodes. The input signals have to travel to the output nodes through the small-world network to achieve high performance.

- Abolition of both the direct connection between $\mathbf{u}(t)$ and $\mathbf{y}(t)$ and the feedback connection between $\mathbf{y}(t)$ and $\mathbf{x}(t)$.

Computing the inverse matrix in Eq. (3) is most expensive. The learning cost of this model is significantly lower than the previous models or standard ESNs as it can reduce the size of the output matrix $T \times N_{\rm out}$.

B. Small-world reservoir

We embed the small-world topology into the reservoir weight matrix, $\mathbf{W}$, based on the Watts and Strogatz method [3]. At first, we construct a regular network of N nodes, where the nodes are arranged in a ring shaped pattern (see Fig. 1 (a)) and each node is connected to its neighboring (previous and next) E nodes. The total number of the connections is $E \times N$, and thus, $\mathbf{W}$ (the $N \times N$ matrix) is sparse. Each directional connection is then rewired with a probability p to another randomly selected node. These connection weights are sampled from an uniform distribution, which constitutes an initial weight matrix $\mathbf{W}_0$. The matrix is scaled by its largest absolute eigenvalue (spectral radius) $|\lambda_{\rm max}|$:

$$\mathbf{W} = \alpha \frac{\mathbf{W}_0}{|\lambda_{\rm max}|}, \tag{4}$$

where α is a scaling constant. This scaling enables the network to obtain the "echo state property" that can sustain appropriate reservoir dynamics [14]. In the standard ESNs, the reservoir possesses the echo state property when its spectral radius satisfies $|\lambda_{\rm max}| < 1$.

III. Experimental configuration

A. Network conditions

We assume the number of reservoir nodes, N, is 1000, and the number of connections for each node, E is 10. The input weights in $\mathbf{W}_{\rm in}$ are drawn from an uniform distribution over $[-0.5, 0.5]$. The scaling constant α in Eq. (4) is set to 1.25. These parameters are fixed in all our experiments. The most important parameter is the number of I/O reservoir nodes, $N_{\rm in}$ and $N_{\rm out}$. We examined the model in the cases where the numbers of the I/O nodes $N_{\rm in} = N_{\rm out}$ = 1000, 500, 300, 100, and 50. When $N_{\rm in} = N_{\rm out} = 1000$, the model is identical to the previous small-world ESNs. In contrast, the model where $N_{\rm in} = N_{\rm out} = 50$ is the most economical.

We tested three kinds of the networks:

- *Fully connected ESN (for reference):* The reservoir nodes are fully connected to each other. The weights in $\mathbf{W}_0$ are sampled from an uniform distribution over $[-0.5, 0.5]$ and do not conform to the small-world topology.
- *Small-world ESN with random I/O nodes:* The reservoir forms a small-world topology. The I/O nodes are randomly selected and can overlap.
- *Small-world ESN with segregated I/O nodes:* The reservoir forms a small-world topology. The $N_{\rm in}$ input nodes are selected in a line on the ring shape, and the $N_{\rm out}$ output nodes are positioned on the opposite side of the input nodes as shown in Fig. 2.

In the cases of the small-world ESNs, the rewiring probability p is set to 0 (regular), 0.001, 0.01, 0.1, 0.2, 0.5, or 1 (random).

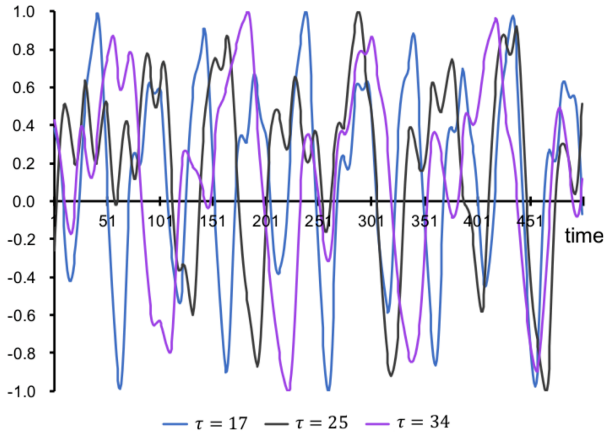


Fig. 3. The Mackey-Glass time-series with $\tau = 17$, 25, and 34, represented by the blue, gray and purple curves, respectively.

B. Data set and evaluation

Our model learned the Mackey-Glass (MG) system [20] that is expressed as a nonlinear time delay differential equation:

$$\frac{{\rm d}x}{{\rm d}t} = \frac{\beta x(t-\tau)}{1 + x(t-\tau)^n} - \gamma x(t), \tag{5}$$

where x denotes the state and τ is the time delay. We set that $\gamma = 0.1$, $\beta = 0.2$, and $n = 10$. Since the current state depends on the state that was recorded τ time steps earlier, a learner of the time-series has to store the input for τ time steps. The time-series with sufficiently large τ behaves chaotically, and therefore, it has often been used as a benchmark task of time-series prediction (e.g., [21]). We solved Eq. (5) using the fourth order Runge–Kutta method. We fed the state into the model as the one-dimensional input $\mathbf{u}(t) \equiv x(t)$ (i.e., $K = 1$), and presented the next state $x(t+1)$ as the one-dimensional desired output $\mathbf{y}_{\rm d}(t) \equiv x(t+1)$ (i.e., $M = 1$).

We prepared three time-series with $\tau = 17$, 25, and 34. Each time-series was scaled between -1 and 1. Fig. 3 shows these time-series between 1 and 500 data points. The reservoir was driven for 4100 data points as it was fed the time-series inputs. The reservoir states for the first 100 data points were discarded taking into consideration the burn-in period. The data points from 101 to 2100 and from 2101 to 4100 were used for training and testing the model, respectively. We evaluated the model by the mean squared error (MSE) between the output and the true MG time-series in the test period. We ran the model 100 times, independently, and measured the median of their MSEs in each case because the MSEs in cases of the unsuccessful learning were extremely large.

We analyzed the basic properties of the reservoir networks in terms of the echo state property and small-worldness. The echo state property is measured by the spectral radius $|\lambda_{\rm max}|$. In general, the reservoir works well if the value is below one or close to one [14]. The small-worldness is defined by the shortest path length, L, and the clustering coefficient, C. These are scaled by their counterparts in the regular network, L_0 and C_0, and the ratio $(C/C_0)/(L/L_0)$ indicates small-worldness.

TABLE I
NETWORK PROPERTIES FOR THE ECHO STATE AND SMALL-WORLDNESS.

Network	$\lvert\lambda_{\max}\rvert$	L/L_0	C/C_0	$(C/C_0)/(L/L_0)$
Fully connected	0.93	–	–	–
$p = 0$ (regular)	1.24	1.00	1.00	1.00
$p = 0.001$	1.24	0.56	1.00	1.81
$p = 0.01$	1.24	0.18	0.98	5.55
$p = 0.1$	1.18	0.09	0.74	8.38
$p = 0.2$	1.12	0.08	0.53	6.46
$p = 0.5$	0.97	0.07	0.14	2.08
$p = 1$ (random)	0.93	0.07	0.02	0.23

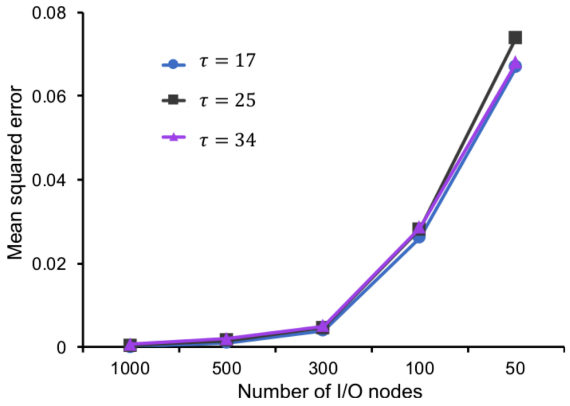


Fig. 4. The mean squared errors of the fully connected echo state networks. They learned the Mackey-Glass equation with $\tau = 17$, 25, or 34, represented by the blue, gray, and purple lines, respectively.

IV. RESULT

A. Network property

The network properties are summarized in Table I. The spectral radiuses of the fully connected network and the random networks ($p = 1$ and 0.5) were less than 1, while those of the regular and small-world networks were greater than 1. This classical theory predicts that the performances of the fully connected and random networks are superior to those of the regular and small-world networks within a standard configuration. The network with $p = 0.1$ had the best small-worldness, i.e., a low L and a high C.

B. Fully connected ESN

Fig. 4 shows the MSEs of the fully connected ESNs for the MG time-series with $\tau = 17$, 25, and 34. The MSEs were very small if the number of the I/O nodes (N_{in} and N_{out}) was 1000, where all nodes could receive input and emit output. However, the MSEs increased as the number of the I/O nodes decreased. This tendency did not depend on the time delay τ in the MG. This result clearly highlights the weakness of the fully connected ESNs that require a sufficiently high number of I/O nodes.

C. Small-world ESN with random I/O nodes

Fig. 5 shows the MSEs of the small-world ESNs where the I/O nodes were randomly selected and could overlap. Fig. 5 (a), (b), and (c) indicate the results for the MG time-series with $\tau = 17$, 25, and 34, respectively, however, they show similar MSE landscapes. The MSEs were very small for random networks ($p = 1$) with many I/O nodes. In contrast, their MSEs increased as the number of I/O nodes decreased. This result is similar to the results of the fully connected random networks shown in Fig. 4. Interestingly, this relation was reversed in the regular network ($p = 0$); the MSEs decreased as the number of the I/O nodes decreased. Regular networks failed to learn of the MG when they had many I/O nodes. The small-world networks ($p = 0.1$) showed an intermediate performance between random and regular networks. With a moderate number of I/O nodes, i.e., between 100 and 300, the small-world networks were superior to the random and regular networks.

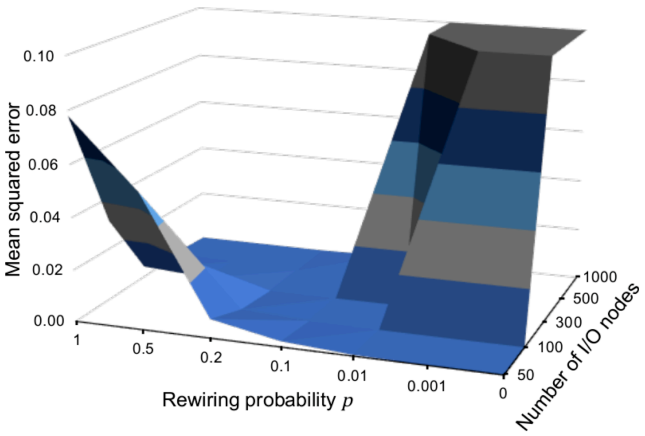
D. Small-world ESN with segregated I/O nodes

Fig. 6 shows the MSEs of those small-world networks where the I/O nodes were fully segregated. These figures provide results for the MG time-series with τ = (a) 17, (b) 25, and (c) 34, respectively. As seen earlier, there were not many significant differences between the time delay τ. The MSE curve in the case of random networks ($p = 1$) is similar to those of the fully-connected networks (refer to Fig. 4) and the random networks with randomly selected I/O nodes (refer to Fig. 5). An important characteristic appeared in the MSEs of regular networks ($p = 0$). Their MSEs remained large even when the number of I/O nodes was 50, which was different from the previous results shown in Fig. 5. Consequently, small-world networks ($p = 0.1$) showed optimal performance when number of the I/O nodes was small.

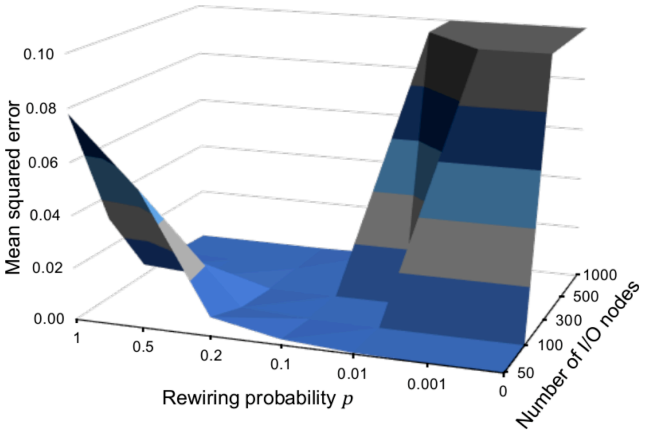
V. DISCUSSION

We demonstrated the advantage of the small-world topology in an economical configuration, i.e., small number of I/O nodes. The results shown in Fig. 5 suggest that the cluster structures in regular networks may be important for storing information that can help predict the future input (refer to Table I). However, the performance declined in the case of a regular network with many I/O nodes and this was probably because the memorized information was easily broken by the external input. In addition, the memory function did not work when the I/O nodes were separated (refer to Fig. 6). I/O segregation forced the network to transfer the input signals to the output nodes, which requires a low shortest pass length between these nodes. Failure to learn in regular networks is caused by their high shortest pass length. In contrast, the small-world network has both a low shortest pass length and a high clustering coefficient (refer to Table I). The capabilities to memorize and transfer information in the small-world networks might improve the learning performances of the ESNs. It is to be noted that the spectral radius does not explain this tendency. We propose that the shortest path length and clustering coefficient are useful parameters to explain the performance of the ESN, especially in an economical configuration.

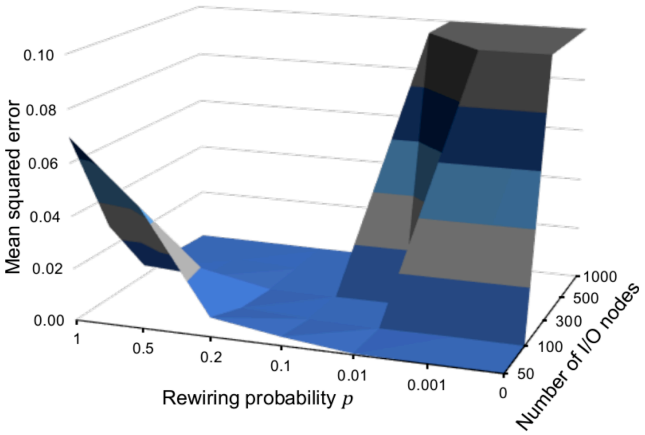
To verify the hypothetical mechanism mentioned above, we should analyze in depth the information flow in a small-world



(a) $\tau = 17$

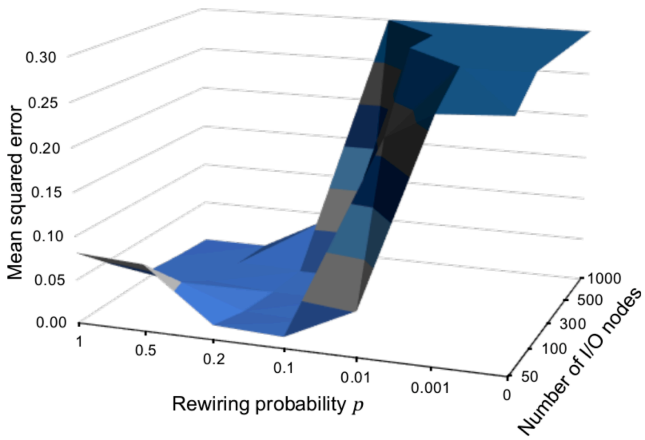


(b) $\tau = 25$

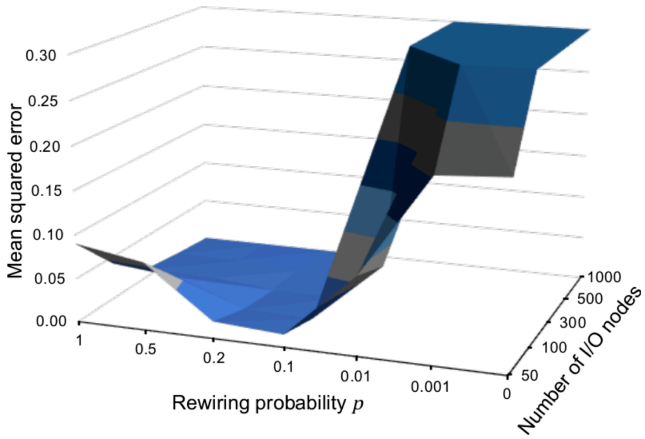


(c) $\tau = 34$

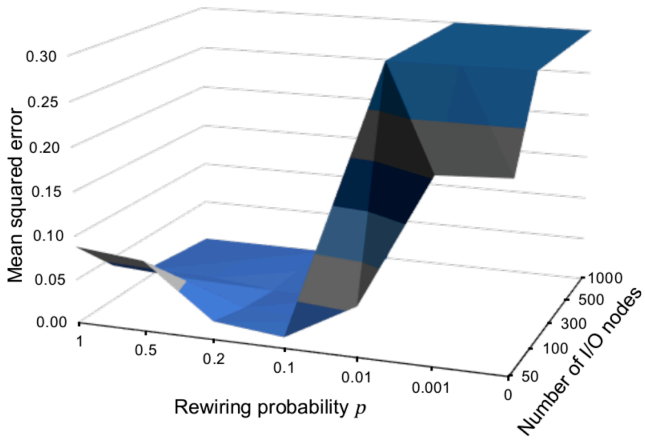
Fig. 5. The mean squared errors of the small-world echo state networks where the reservoir nodes receiving external input or sending output were randomly selected. The axis in the depth direction indicates the number of these input/output nodes, which was limited to smaller than those of all reservoir nodes. The networks learned the Mackey-Glass equation with $\tau =$ (a) 17, (b) 25, or (c) 34.



(a) $\tau = 17$



(b) $\tau = 25$



(c) $\tau = 34$

Fig. 6. The mean squared errors of the small-world echo state networks where the reservoir nodes receiving input were selected in line on the ring of the nodes, and the nodes sending output were arranged on the opposite side of the input nodes (refer to Fig. 2). The axis in the depth direction indicates the number of these input/output nodes that are smaller than those of all reservoir nodes. The networks learned the Mackey-Glass equation with $\tau =$ (a) 17, (b) 25, or (c) 34.

reservoir. Boedecker et al. [22] quantified computational capabilities of the randomly connected ESNs using information-theoretical measures. Eventually, we would like to establish a theory for relationship between the graph-theoretical parameters (e.g., the shortest path length and clustering coefficient) of neural networks and the information-theoretical measures of their neural dynamics.

The ESNs with the small-world topology displayed high performance under two constraints: reduction as well as segregation of the I/O nodes. Inversely, the small-world topology may be a result of evolutionary optimization under these constraints. The first constraint is relative to the wiring cost, i.e., minimization of total axonal length. The wiring cost has been considered as an important evolutionary pressure [23], [24]. Harter [24] reported the emergence of a small-world topology in the brain network of a robot as a result of the optimization of its connectivity by an evolutionary algorithm considering the wiring cost and task performance. We propose, that in addition to the wiring cost, the second constraint, i.e., I/O segregation, be the key to the emergence of the small-world topology. Even under this constraint, the small-world topology enables the network to rapidly and efficiently transfer input information to output regions.

Characteristic topologies of brain networks are not only small-world [4], [6], [7] but also scale-free [25] and rich-club [26]. These topologies emphasize the existence of hub cortical regions that have many connections with other regions. The hub nodes seem to have a great impact on neural dynamics. We can understand more about cortical information processing if our constructive approach is able to clarify the role of the hub organization.

VI. Conclusion

We found that the small-world topology is advantageous in neural information processing of the ESNs although the conventional studies [17], [18] have not shown this. The key settings were reduction and segregation of I/O nodes. The small-world ESNs exhibited better performance in time-series prediction than the random and regular ESNs when the number of I/O nodes was much smaller than those of the overall nodes, and when the I/O nodes were fully segregated. This high performance might relay on the high memory capability and fast information transfer of the small-world networks. This finding leads to better understanding of cortical information processing, an evolutionary origin of brain networks, as well as more economical and efficient ESNs.

Acknowledgment

This research and development work was supported by the MIC.

References

[1] S. Achard and E. Bullmore, "Efficiency and cost of economical brain functional networks," *PLoS Computational Biology*, vol. 3, no. 2, p. e17, 2007.

[2] E. Bullmore and O. Sporns, "The economy of brain network organization," *Nature Reviews Neuroscience*, vol. 13, no. 5, pp. 336–349, 2012.

[3] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[4] S. Achard, R. Salvador, B. Whitcher, J. Suckling, and E. Bullmore, "A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs," *Journal of Neuroscience*, vol. 26, no. 1, pp. 63–72, 2006.

[5] M. J. Vaessen, P. A. M. Hofman, H. N. Tijssen, A. P. Aldenkamp, J. F. A. Jansen, and W. H. Backes, "The effect and reproducibility of different clinical DTI gradient sets on small world brain connectivity measures," *Neuroimage*, vol. 51, no. 3, pp. 1106–1116, 2010.

[6] O. Sporns and J. D. Zwi, "The small world of the cerebral cortex," *Neuroinformatics*, vol. 2, no. 2, pp. 145–162, 2004.

[7] D. S. Bassett and E. Bullmore, "Small-world brain networks," *Neuroscientist*, vol. 12, no. 6, pp. 512–523, 2006.

[8] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nature Reviews Neuroscience*, vol. 10, no. 3, pp. 186–198, 2009.

[9] D. Simard, L. Nadeau, and H. Kröger, "Fastest learning in small-world neural networks," *Physics Letters A*, vol. 336, no. 1, pp. 8–15, 2005.

[10] L. Xiaohu, L. Xiaoling, Z. Jinhua, Z. Yulin, and L. Maolin, "A new multilayer feedforward small-world neural network with its performances on function approximation," in *Proceedings of the IEEE International Conference on Computer Science and Automation Engineering*, 2011, pp. 353–357.

[11] O. Erkaymaz, M. Özer, and N. Yumuşak, "Performance analysis of a feed-forward artificial neural network with small-world topology," *Procedia Technology*, vol. 1, pp. 291–296, 2012.

[12] J. W. Bohland and A. A. Minai, "Efficient associative memory using small-world architecture," *Neurocomputing*, vol. 38–40, pp. 489–496, 2001.

[13] L. G. Morelli, G. Abramson, and M. N. Kuperman, "Associative memory on a small-world neural network," *European Physical Journal B – Condensed Matter and Complex Systems*, vol. 38, no. 3, pp. 495–500, 2004.

[14] H. Jaeger, "The "echo state" approach to analysing and training recurrent neural networks," German National Research Center for Information Technology, Tech. Rep., 2001.

[15] W. Maass, T. Natschläger, and H. Markram, "Real-time computing without stable states: A new framework for neural computation based on perturbations," *Neural Computation*, vol. 14, no. 11, pp. 2531–2560, 2002.

[16] P. F. Dominey, "Complex sensory-motor sequence learning based on recurrent state representation and reinforcement learning," *Biological Cybernetics*, vol. 73, no. 3, pp. 265–274, 1995.

[17] B. Liebald, "Exploration of effects of different network topologies on the ESN signal crosscorrelation matrix spectrum," Bachelor thesis, Jacobs University Bremen, 2004.

[18] A. A. Rad, "Dynamical networks (miniporject) effect of topology of the reservoir on performance of echo state networks," http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.94.9424, 2008.

[19] Z. Deng and Y. Zhang, "Collective behavior of a small-world recurrent neural system with scale-free distribution," *IEEE Transactions on Neural Networks*, vol. 18, no. 5, pp. 1364–1375, 2007.

[20] M. C. Mackey and L. Glass, "Oscillation and chaos in physiological control systems," *Science*, vol. 197, no. 4300, pp. 287–289, 1977.

[21] K. S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Transactions on Neural Networks*, vol. 1, no. 1, pp. 4–27, 1990.

[22] J. Boedecker, O. Obst, J. T. Lizier, N. M. Mayer, and M. Asada, "Information processing in echo state networks at the edge of chaos," *Theory in Biosciences*, vol. 131, no. 3, pp. 205–213, 2012.

[23] D. B. Chklovskii, T. Schikorski, and C. F. Stevens, "Wiring optimization in cortical circuits," *Neuron*, vol. 34, no. 3, pp. 341–347, 2002.

[24] D. Harter, "Functional and physical constraints for evolving small-world structure in embodied networks," in *Proceedings of the IEEE International Joint Conference on Neural Networks*, 2011, pp. 2357–2362.

[25] V. M. Eguiluz, D. R. Chialvo, G. A. Cecchi, M. Baliki, and A. V. Apkarian, "Scale-free brain functional networks," *Physical Review Letters*, vol. 94, no. 1, p. 018102, 2005.

[26] M. P. Van Den Heuvel and O. Sporns, "Rich-club organization of the human connectome," *Journal of Neuroscience*, vol. 31, no. 44, pp. 15 775–15 786, 2011.