

SORN: a Self-organizing Recurrent Neural Network

Andreea Lazar^{1*}, Gordon Pipa^{1,2}, Jochen Triesch¹

¹Frankfurt Institute of Advanced Studies, Johann Wolfgang Goethe University, Frankfurt am Main, Germany

²Department of Neurophysiology, Max Planck Institute for Brain Research, Frankfurt am Main, Germany

Correspondence:

Andreea Lazar
Frankfurt Institute for Advanced Studies (FIAS),
Ruth-Moufang-Str.1, 60438 Frankfurt am Main, Germany
lazar@fias.uni-frankfurt.de

Running Title:

SORN: a Self-organizing Recurrent Network

Supplemental Data

The choice of threshold values for excitatory (T_{max}^E) and inhibitory units (T_{max}^I) plays an important role in determining the dynamics of a random reservoir network. In the following experiments we consider 10 networks with $N^E = 100$ (Fig. 1) or $N^E = 200$ (Fig. 2) for 15 different settings of T_{max}^E and T_{max}^I . We compare static reservoirs to SORNs following 50,000 steps of plasticity in the presence of a temporally patterned input (counting task with $n = 8$). We simulate each network for 5,000 time steps of activity (frozen weights and thresholds for SORNs) and monitor the following aspects of the network dynamics:

- **Rate H_0** - defined as the mean fraction of firing neurons per unit of time.
- **Inactive neurons** - defined as the fraction of units which are silent during simulation.
- **Criticality of network dynamics** - assessed through a perturbation analysis. For every state $x(t)$, we perturb the activation of a randomly chosen excitatory neuron (from active to inactive or from inactive to active) creating an altered state $\tilde{x}(t)$. The Hamming distance between $x(t)$ and its perturbed version $\tilde{x}(t)$ is one ($d(t) = 1$). We calculate the successor states of $x(t)$ and $\tilde{x}(t)$ by applying (1) and obtain $x(t+1)$ and $\tilde{x}(t+1)$ with the Hamming distance $d(t+1)$. If the average distance $\bar{d}(t+1) > 1$ the network amplifies perturbations and is in a supercritical regime. If $\bar{d}(t+1) < 1$ the network has self-correcting properties and is in a subcritical dynamical regime. When $\bar{d}(t+1) \approx 1$ the dynamics is said to be on the “edge of chaos”.
- **One-step prediction performance** - A readout is trained in a supervised fashion to predict the next input ($U(t)$) based on the network’s internal state ($x'(t)$) after presentation of the preceding letter ($U(t-1)$). We use the Moore–Penrose pseudoinverse method that minimizes the squared difference between the output of the readout neurons and the target output value. The quality of the readout (test performance) is assessed on a second sample of 5000 steps of activity using an independent input sequence.

The results prompt us to the following observations:

- For static reservoirs, small network rates correspond to high fractions of inactive neurons. For SORNs, IP successfully spreads the network activity across all neurons.
- Static reservoirs with critical dynamics score highest in prediction performance when compared to static networks in supercritical or subcritical dynamical regimes. In contrast, SORNs settle into a subcritical regime and exhibit a nearly optimal performance.

This analysis allows us to identify the static networks with critical dynamics and high prediction performance. In our experiments, we challenge SORNs to outperform these highly tuned static reservoirs. A less careful choice of initial threshold settings results into a similar or larger performance advantage for SORNs compared to static networks.

$N^E = 100$

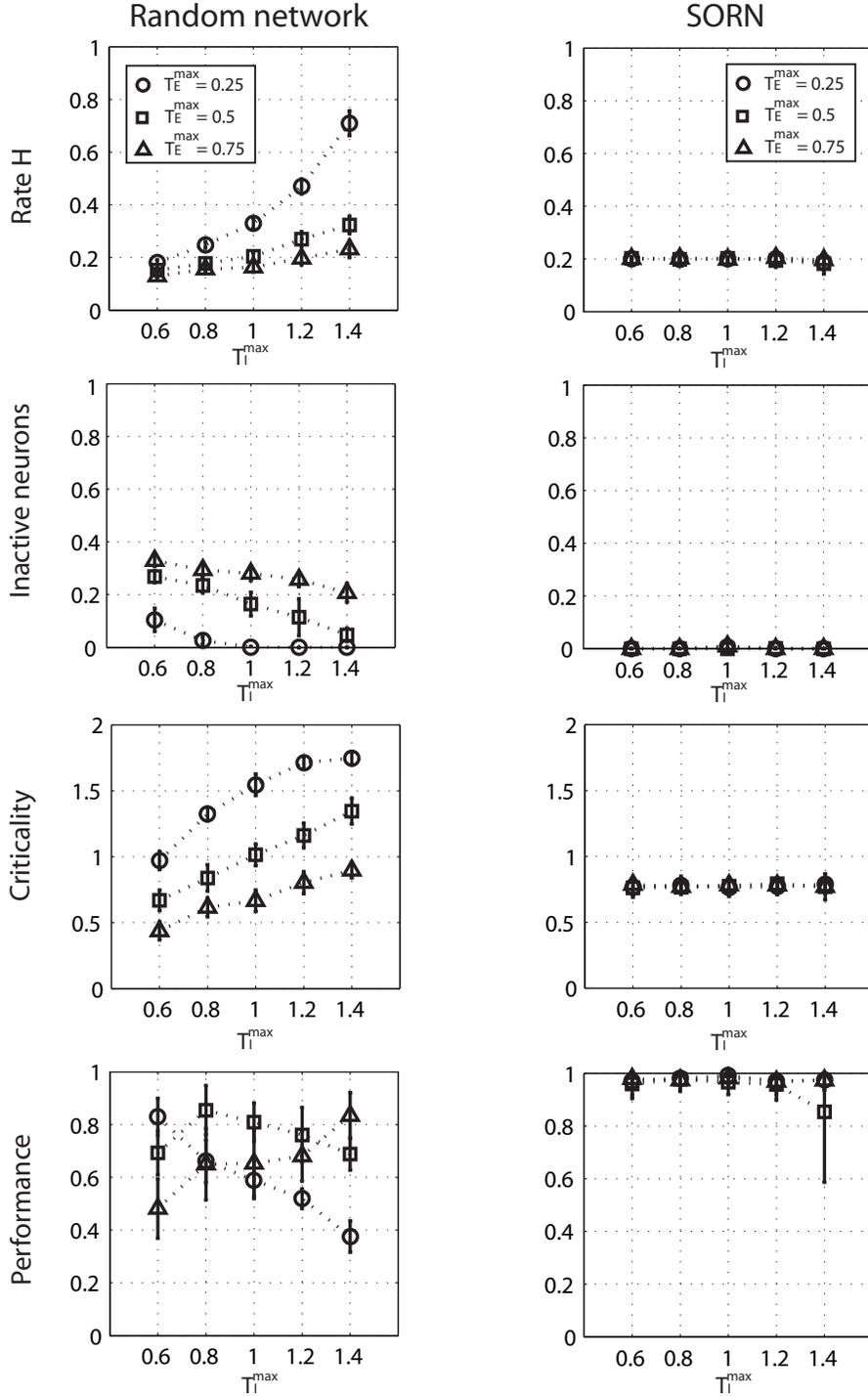


Figure 1: Influence of different initial thresholds on network dynamics: static reservoirs with $N^E = 100$, $N^U = 10\% \times N^E$, $\lambda^W = 10$ are compared to their corresponding SORNs with $H_{IP} = 20\% \times N^E$. Error bars indicate standard error.

$N^E = 200$

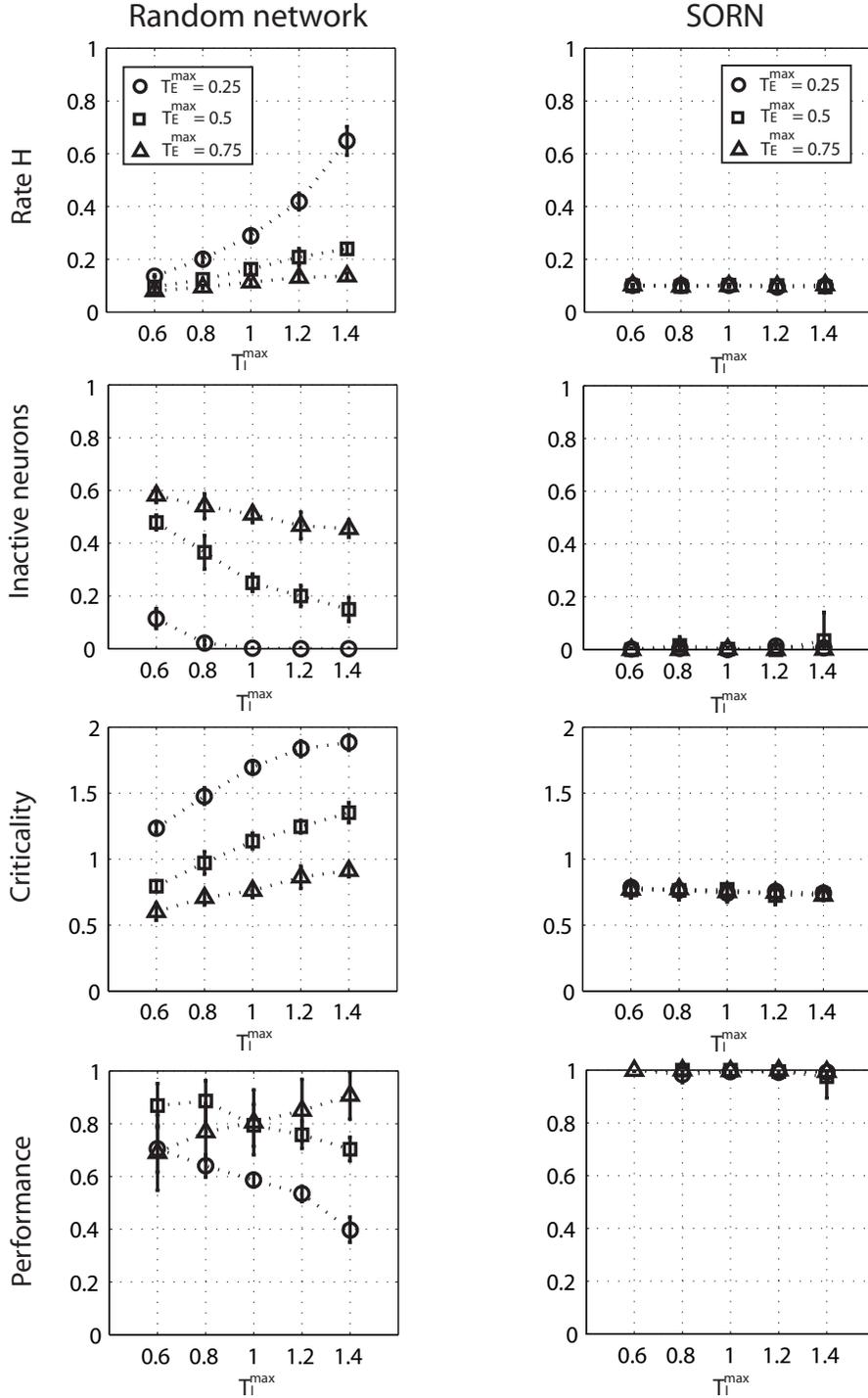


Figure 2: Influence of different initial thresholds on network dynamics: static reservoirs with $N^E = 200$, $N^U = 10\% \times N^E$, $\lambda^W = 10$ are compared to their corresponding SORNs with $H_{IP} = 20\% \times N^E$. Error bars indicate standard error.