

A CHARACTERIZATION OF THE EDGE OF CRITICALITY IN BINARY ECHO STATE NETWORKS

Pietro Verzelli¹, Lorenzo Livi², Cesare Alippi^{1,3}

¹Università della Svizzera Italiana, Lugano,Switzerland. ² University of Exeter, Exeter,United Kingdom. ³ Politecnico di Milano, Milano,Italy.

ABSTRACT

Echo State Networks (ESNs) are simplified recurrent neural network models composed of a reservoir and a linear, trainable readout layer. The reservoir is tunable by some hyper-parameters that control the network behaviour. ESNs are known to be effective in solving tasks when configured on a region in (hyper-)parameter space called *Edge of Criticality* (EoC), where the system is maximally sensitive to perturbations hence affecting its behaviour. In this paper, we propose binary ESNs, which are architecturally equivalent to standard ESNs but consider binary activation functions and binary recurrent weights. For these networks, we derive a closed-form expression for the EoC in the autonomous case and perform simulations in order to assess their behavior in the case of noisy neurons and in the presence of a signal. We propose a theoretical explanation for the fact that the variance of the input plays a major role in characterizing the EoC.

Index Terms— Reservoir computing; Binarization; Random Boolean networks; Edge of Criticality.

1. INTRODUCTION

Echo State Networks (ESNs) [6] maximize predictive performance on the Edge of Criticality or Edge of Chaos (EoC), which is a region in parameter space where the system is maximally sensitive to perturbations [9]. To date, a complete theoretical understanding for the EoC is missing for input-driven ESNs without mean-field assumptions. This performance maximization at the EoC is known to be common to many dynamical systems, even really simple ones [8]. This property has lead many researcher to study these models in order to explore the main features associated with the transition to chaos. In this direction, Random Boolean Networks (RBNs) [7] are well studied networks. Their chaotic behavior is well understood ([3, 4, 14]), though their applicability is scarce due to the that they only input binary signals.

In this paper, we propose binary ESNs (bESNs). The architecture is equivalent to standard ESNs but simplified as they consider binary activation functions and binary weights for the recurrent connections. To the best of our knowledge, this architecture has never been investigated before. bESNs

share some similarities also with a particular form of RBNs, called Random Threshold Networks [13] (that are based on the same idea of summing the input of neurons). We derive a closed-form expression to determine the EoC in autonomous bESNs (i.e., the network is not driven by signals) and perform simulations to assess the behavior of bESNs both in the autonomous and non-autonomous case. We experimentally assess the quality of our theoretical prediction regarding the onset of chaos in bESNs in the autonomous case. Results show perfect agreement with the theory. Then, in order to assess the network stability, we analyze the impact of noise on the neuron outputs on the onset of chaos. Our findings suggests that the chaotic region expands linearly with the noise intensity when the mean degree (i.e., the average number of links a neuron has) is high enough. We also study the EoC for bESNs driven by (continuous) signals, discussing how our findings could be generalized considering the signal gain as an hyperparameter. This work sets also in the context of reducing model complexity, in which binarization plays an important role both for theoretical aspects [1] and applied perspectives [2], since it would considerably reduce required hardware resources and speed-up training algorithms.

2. BACKGROUND MATERIAL

2.1. Echo State Networks

An ESN is a discrete-time non-linear system with feedback, whose model reads:

$$\mathbf{x}[n+1] = f(\mathbf{W}\mathbf{x}[n] + \mathbf{W}_i^r \mathbf{u}[n+1] + \mathbf{W}_o^r \mathbf{y}[n]) \quad (1)$$

$$\mathbf{y}[n+1] = g(\mathbf{W}_i^o \mathbf{u}[n+1] + \mathbf{W}_r^o \mathbf{x}[n+1]) \quad (2)$$

An ESN consists of a *reservoir* of N neurons characterized by a non-linear transfer function $f(\cdot)$. At time n the network is driven by the input $\mathbf{u}[n] \in \mathbb{R}^{N_i}$ and produces an output $\mathbf{y}[n] \in \mathbb{R}^{N_o}$, N_i and N_o being the dimensions of the input and output vectors, respectively.

The weight matrices $\mathbf{W} \in \mathbb{R}^{N \times N}$ (reservoir internal connections), $\mathbf{W}_i^r \in \mathbb{R}^{N \times N_i}$ (input-to-reservoir connections) and $\mathbf{W}_o^r \in \mathbb{R}^{N \times N_o}$ (output-to-reservoir feedback connections) contain values in the $[-1, 1]$ interval drawn from a uniform (or sometimes Gaussian) distribution. The output weight

matrices $\mathbf{W}_i^o \in \mathbb{R}^{N_o \times N_i}$ and $\mathbf{W}_r^o \in \mathbb{R}^{N_o \times N_r}$, connecting reservoir and input to the output, represent the readout layer of the network. Activation functions $f(\cdot)$ and $g(\cdot)$ (applied component-wise) are typically implemented as a sigmoidal (tanh) and identity function, respectively. Training requires solving a regularized least-square problem [6].

Various empirical results suggest that ESNs achieve the highest expressive power, i.e., the ability to provide optimal performances, exactly when configured on the edge of the transition between a ordered and chaotic regime (e.g., see [6, 9–12, 15]). Once the network operates on the edge -or in proximity to - it achieves the highest memory capacity (storage of past information) and accuracy prediction, compatible with the network architecture. For determining the edge of chaos, one usually resorts to computing the *maximum Lyapunov exponent* [5] or identify parameter configuration maximizing the *Fisher information* [10].

2.2. Random Boolean Networks

RBNs were first proposed in [7] as a model for the gene regulatory mechanism. The model consists of N variables $\sigma_i \in \{0, 1\}$ – sometimes called *spins* – whose time-evolution is given by $\sigma_i[n+1] = f_i(\sigma_{i_1}[n], \sigma_{i_2}[n], \dots, \sigma_{i_K}[n])$, where each $f_i \in \{0, 1\}$ is a Boolean function of K variables, representing the K incoming links to the i -th element of the net. There exists 2^{2^K} possible Boolean functions of K variables. The output of f_i is randomly chosen to be 1 with probability r and 0 with probability $1-r$, so that usually one refers to r as the *bias* of the network.

RBNs show two distinct regimes, depending on both the value of K and r : a phase in which the network assumes a *chaotic* behaviour and a phase (sometimes called *frozen*) in which the network rapidly collapses to a stable state. In [3], the authors justify this behavior by studying the evolution of the (Hamming) distance of two (initially different) configurations. There, they derive the following formula for the onset of the chaos: $K > K_c = \frac{1}{2r(1-r)}$.

3. BINARY ECHO STATE NETWORK

In this section, we introduce bESNs and study the dynamics of a reservoir similar to (1) constituted of binary neurons $x_i \in \{-1, +1\}$ for $i = 1, 2, \dots, N$ and binary weights $W_{ij} \in \{-1, 0, +1\}$ for $i, j = 1, 2, \dots, N$ (the zero value accounts for the fact that two neurons may not be linked). For simplicity, we will not consider feedback connection (i.e., $\mathbf{W}_o^r = 0$). The bESNs system model simplifies as:

$$x_i[n+1] = \text{sgn}(S_i[n]), \quad (3)$$

$$S_i[n] := \sum_{j=1}^N W_{ij} x_j[n] + u[t]. \quad (4)$$

$u[t]$ is the input signal, which we consider to be unfiltered ($(\mathbf{W}_i^r) = \mathbf{1}$, i.e., the all-ones vector), for simplicity. When $u[n] = 0$ for every n (i.e., there is no input), we say that the system is *autonomous*. The study of the autonomous system plays an important role, since it allows us to investigate analytically the network dynamics and its properties. Reservoir connections $\mathbf{W} = (W_{ij})$ are instantiated according to the Erdős–Rényi model where each link W_{ij} is created with probability α . If the link is generated, the weight value is set to 1 with probability p or -1 with probability $1-p$.

The proposed bESN model is controlled by three hyperparameters: (1) N , the *number of neurons* in the network; (2) $\langle k \rangle := \alpha N$, the *mean degree* of the network; (3) $d := p - \frac{1}{2}$, the *asymmetry* in the weights values. These hyperparameters are related to α and p , although they are easier to understand: in fact, $\langle k \rangle$ has a natural interpretation in terms of mean neuron degree that does not depend on the network size N . The choice of using d is due to the symmetry of the model around the zero value and to the fact that, with this choice, a positive (negative) value of the hyperparameter accounts for majority of positive (negative) weights. Note that $\langle k \rangle$ can vary *continuously* from 0 to N and $d \in (-\frac{1}{2}, \frac{1}{2})$. A similar model was proposed in [13], but in their work the weights assume a positive or negative value with equal probability, i.e., their model corresponds to ours in the $p = 0$ case.

3.1. Edge of chaos in binary ESN

Here, we study two networks with the same weight matrix \mathbf{W} , that are in the states $\mathbf{x}[n]$ and $\mathbf{x}'[n]$, the latter refers to the perturbed network and differs only in one neuron whose state is flipped. The goal is here to understand under which conditions the time evolution of the perturbed network differs from the original one, i.e., whether the perturbation will spread and significantly impact the network behavior or not.

For $N \rightarrow \infty$, the fraction of positive-valued neurons is equal to the probability for a neuron of being positive, namely $P_+ = p$, while the fraction of negative neurons is $P_- = q = 1 - p$. By comparing the original network with the perturbed one, the probability that the flipped neuron will have an influence on a neuron connected to it will be given by two terms: the probability that neuron state is positive P_+ multiplied by the probability of switching from positive to negative π_{+-} , plus an analogous terms accounting for the negative part (P_- and π_{-+} respectively). Assuming that $\pi_{+-} = P_- = q$ (i.e., that the probability of turning negative from positive is equal to the probability of being negative) and, analogously, that $\pi_{-+} = P_+ = p$ (that can be seen as a formulation of the *annealed approximation* introduced in [3]), one obtains:

$$P_+ \cdot \pi_{+-} + P_- \cdot \pi_{-+} = pq + qp = 2p(1-p) \quad (5)$$

We now define k_O and k_I as the mean *out-degree* and *in-degree* of a neuron, respectively. Since a single neuron has influence on k_O neurons, the expected number of changes

is given by $2p(1-p)k_O$, to which one has to add the fact that at least one neuron has changed due to the flip. Therefore, if this number is bigger than half of the mean incoming links of a neuron, i.e., $k_I/2$, then the perturbation will dominate the network dynamics and will propagate. Since in an Erdős–Rényi graph $k_I = k_O = \langle k \rangle$, we obtain the following condition for the onset of chaos:

$$1 + 2p(1-p)\langle k \rangle > \frac{\langle k \rangle}{2} \quad (6)$$

which using $d := p - 1/2$ can be rewritten as:

$$\langle k \rangle < k_c := \frac{1}{2d^2} \quad (7)$$

Note that the mean degree $\langle k \rangle$ in (7) plays a “stabilizing” role (i.e., the higher the degree, the larger the magnitude of d required for chaos), as opposed to RBNs, where increasing the mean degree leads towards a chaotic region.

4. EXPERIMENTS

4.1. Edge of chaos

In order to assess the agreement between the prediction given by Eq. 7 and experimental results, we conducted an exploration of the parameter space. Here, we exploit the fact that our neurons assume binary states only and consider their *Shannon entropy* H as an indicator for the transition to chaos. The entropy H was computed considering a time average \bar{H} ,

$$\bar{H} := \frac{1}{T - t_0} \sum_{n=t_0}^T H(\mathbf{x}[n]) \quad (8)$$

where the entropy of a configuration \mathbf{x} is estimated as $H(\mathbf{x}) := -\rho(\mathbf{x}) \log(\rho(\mathbf{x})) - (1 - \rho(\mathbf{x})) \log(1 - \rho(\mathbf{x}))$, in which $\rho(\mathbf{x})$ is the number of neurons whose state is +1 and $(1 - \rho(\mathbf{x}))$ the number of neurons with state -1. We expect Eq.(8) to be almost zero in the frozen regime and almost one in the chaotic one, with a sharp region of intermediate values that we consider to be the edge of chaos.

In order to explore the parameter space, we run a series of simulations using a network of $N = 1000$ neurons with different random initial conditions and connection matrices \mathbf{W} , generated using specific values of $\langle k \rangle$ and d . In Eq. 8, we used $T = 300$ time-steps and $t_0 = 100$ accounting for an initial transient from the initial state to a stationary condition. Results are showed in Fig. 1 and demonstrate almost perfect agreement with the theoretical result (7).

4.2. Effects of perturbations on state evolution

In order to assess the effect of chaos on the network behavior, we compare the evolution of a BESN instantiated with weight

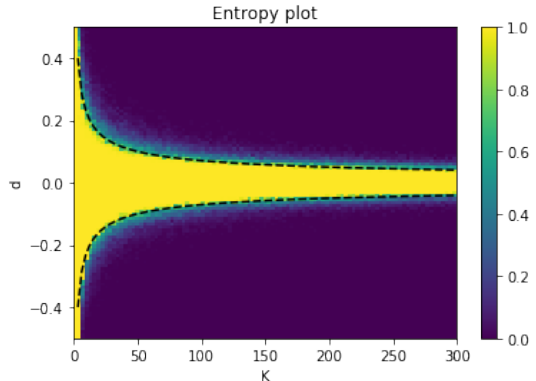


Fig. 1: Values of \bar{H} for different configurations of the asymmetry and the mean degree. The experiment shows a good agreement with the predicted EoC region (dashed line), where we observe an abrupt change of the entropy from 0 to 1.

matrix \mathbf{W} but different initial conditions. Starting from a random initial condition, we generated 50 additional initial conditions by flipping the state of a single neuron (as in Sec. 3.1). Here, an initial condition \mathbf{x}_0 is randomly generated with a biased probability $c = 0.6$ for a neuron to assume a positive value. We let the original and perturbed networks evolve, and take into account the (normalized) Hamming distance, D_H , between trajectories.

Results are summarized in Fig. 2. We observe that, in the ordered phase, perturbations on the initial state have no effect on the network evolution and the Hamming distance of the perturbed trajectory from the original one is zero. As d decreases (i.e., the networks approach the chaotic regime), we observe how the Hamming distance significantly increases, leading to chaos. Note that the maximum value achievable by the (normalized) Hamming distance is 0.5, corresponding to the distance of two random binary vectors (a larger distance would imply a negative correlation). In the same set of figures, we show three additional indicators, called *Energy*, *Activity*, and *Entropy*. The mean *Energy*, defined as $E(\mathbf{x}[n]) := \frac{1}{N} \sum_{i=1}^N x_i[n]$, quantifies the average number of positive and negative neurons. In the frozen phase, the network almost instantly evolves towards values close to 1 (cfr. the role of c , discussed above), and then rapidly decreases to 0, which is the expected value in the chaotic phase. The mean *Activity* of network at time-step n is defined as the (normalized) Hamming distance of the current state w.r.t the previous one, $A(\mathbf{x}[n]) = D_H(\mathbf{x}[n], \mathbf{x}[n-1])$, i.e., the number of neurons that changed their states in one step. As expected, networks operating in chaotic regimes are characterized by an elevated activity. Lastly, we plot the evolution of the *Entropy* (8) over time. As expected from the theory, transitioning to a chaotic regime is signaled by a sharp increase of entropy.

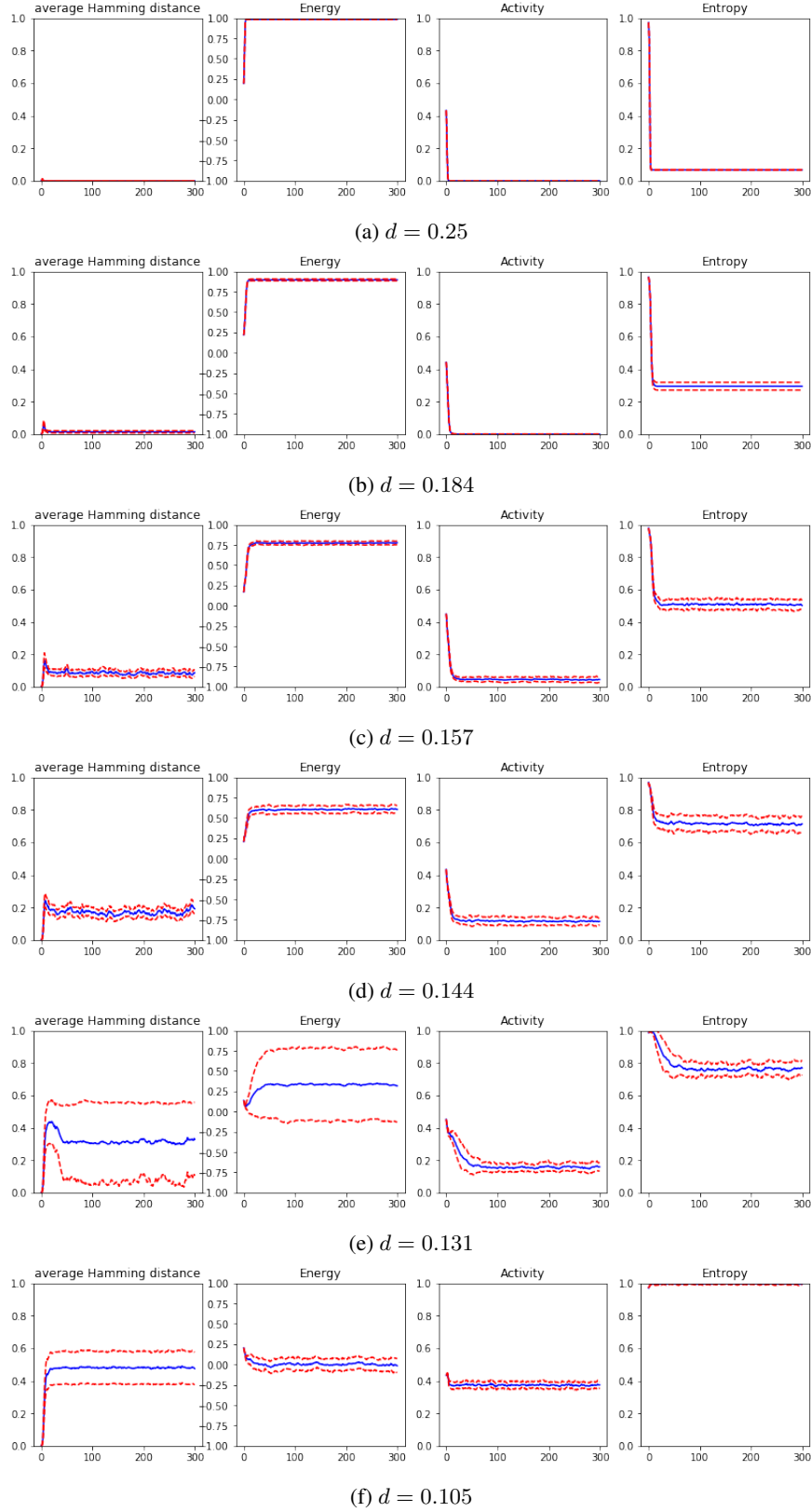


Fig. 2: Mean values of the Hamming distance, Energy, Activity and Entropy of the 50 perturbed networks, with $N = 1000$ and $K = 22$ for selected values of d (see 4.2). The x -axis represents time. The values of the quantities are plotted in blue, while the dashed red lines show the variance. The predicted system should turn chaotic for $d < 1/\sqrt{2\langle k \rangle} \approx 0.15$, according to the theoretical formula.

4.3. Impact of noise in bESN edge of chaos

Here, we study how the EoC is influenced when considering an independent noise term for each neuron, $x_i[n+1] = \text{sgn}(S_i[n] + \nu \cdot \langle k \rangle \cdot \xi_i[n])$, where ν is the *noise gain*, $\xi_i \sim \mathcal{N}(0, 1)$, and S_i is the same as in Eq.(4). The choice of scaling the noise with $\langle k \rangle$ was made to account for the fact that the network stability increases with it, as we discuss below.

To explore the dependency from ν , we ran an experiment where we fixed $\langle k \rangle$ and plotted ν versus d . Results are shown in Fig. 3. We can recognize three regimes: (1) for low noise values, the chaotic region remains almost constant; (2) for intermediate values, the chaotic region linearly expands with the noise intensity up until (3) there is only chaos. We repeated the experiments with different values of $\langle k \rangle$ (not shown) and they all confirm the same linear expansion of the chaotic region (in units of $\langle k \rangle$). To verify this fact for a wider

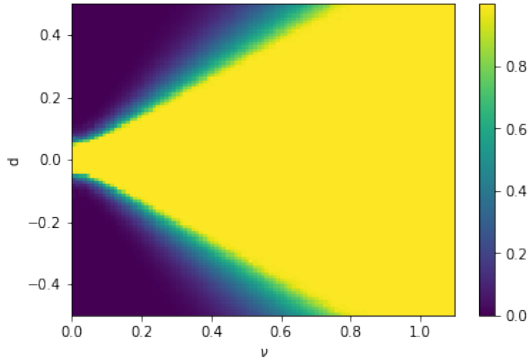


Fig. 3: Values of \bar{H} for different configurations of d and the ν . Note that ν is multiplied by the mean degree, which is here fixed to $\langle k \rangle = 200$.

range of $\langle k \rangle$, we repeated the experiment in Fig. 1 with noise intensity $\nu = 0.1$. It is possible to observe in Fig. 4 how the EoC maintains its shape for lower values of $\langle k \rangle$, while for higher average degrees it deviates from the theoretical prediction and the chaotic region depends on d only. We explain this fact as follows. Neurons can only assume 1 or -1 values. The probability of a neuron having j positive inputs is then $P_k(j) = \binom{k}{j} p^j q^{k-j}$ where k is its in-degree. If we consider that $j = \frac{k+s}{2}$, where s is the value of the *sum of the positive and negative inputs* (whose sign determines the value of the neuron), then we obtain $s = 2j - k$. The expectation of j is $\langle j \rangle = pk$, so that the expectation of s and its variance are:

$$\langle s \rangle = k(p - q) = 2kd \quad (9)$$

$$\langle (s - \langle s \rangle)^2 \rangle = 4kpq = k(1 - 4d^2) \quad (10)$$

Note that these values are related to a single neuron. For a general understanding of the network behavior, one simply uses $\langle k \rangle$ instead of k in the expressions above, so that it is possible to consider $\sigma^2 := \langle k \rangle (1 - 4d^2)$ as a *mean-field variance* of the total inputs to neurons. The impact of the noise

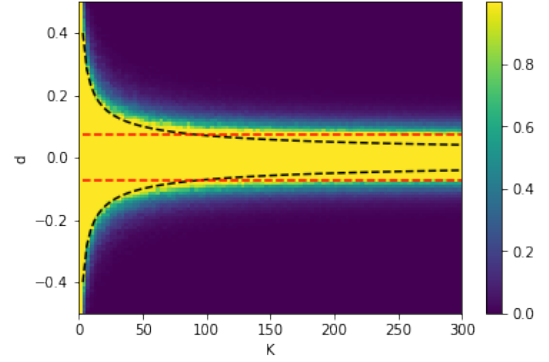


Fig. 4: The same experiment of Fig.1, but with the presence of a noise term with $\nu = 0.1$ (multiplied by the mean degree, so that it increases along the x -axis). Note how for higher degree the chaos region is constant (the predicted value is the red dashed line), deviating from the autonomous-case prediction (red dashed line).

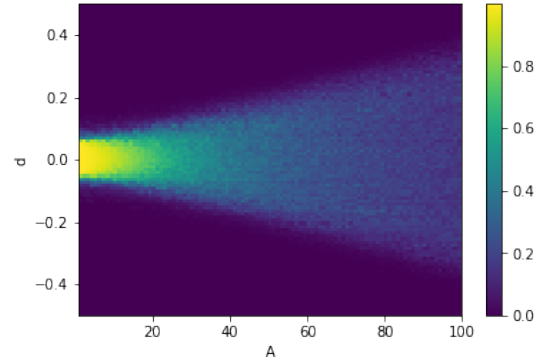


Fig. 5: Network driven by white noise. Values of \bar{H} for different configurations of A and d . The mean degree was fixed to $\langle k \rangle = 150$.

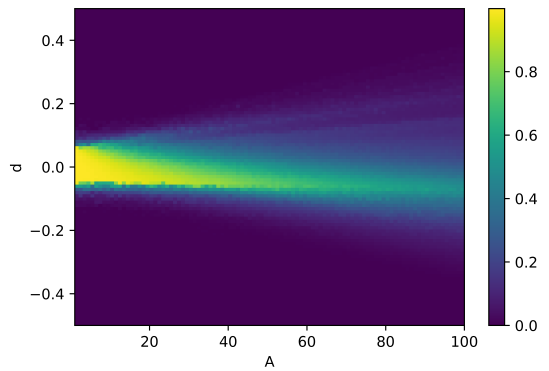


Fig. 6: Network driven with the sum of three sinusoids. Values of \bar{H} for different configurations of A and d . The mean degree was fixed to $\langle k \rangle = 150$.

on the network can then be studied considering the ratio between ν and σ^2 . As previously discussed, the noise expands the chaos region linearly with its ν .

The noise we are considering has a standard deviation $\theta_k = \nu \cdot \langle k \rangle$. This leads us to a formula for the chaotic region which, for $\langle k \rangle \gg 1$, is $|d| < a \cdot \nu + b$. This relation, as shown in Fig. 4, does not depend on $\langle k \rangle$. As such, having Gaussian noise with standard deviation θ , the formula is $|d| < a \cdot \frac{\theta}{\langle k \rangle}$, or in terms of $\langle k \rangle$, we have $\langle k \rangle < k_c^{\text{noise}} := \frac{a\theta}{|d|}$. In our experiments, constant a was determined as $a \approx 0.65$.

4.4. Impact of a signal

As for the noise, the magnitude of the signal should have a major role in the EoC, but this time the chaotic region should reduce instead of expanding, since the signal is known to suppress chaos in certain conditions [11]. The signal introduces a correlation among neurons, which makes the annealed approximation ineffective. We drive the network with the signal as in (4), but we scale $u[n]$ with a *gain* factor A , since we are interested in its usage as an hyperparameter and not in relation with $\langle k \rangle$. We initially feed the network with white noise (note that this is different from what we did in 4.3, since in this case the noise is the same for each neuron): from Fig. 5 one can observe how the chaotic region rapidly shrinks as A increases, but a region with an intermediate value of entropy expands (linearly). This is due to the fact the signal prevents the system from collapsing in a stable state, keeping the entropy above zero.

In Fig.6 we show the results obtained for the normalized sum of three sines with incommensurable frequencies (repeated also with different numbers of sinudoids, not shown). Again we note how the chaotic region shrinks as A increases, with the appearance of the region characterized by intermediate entropy values which, instead, expands.

5. CONCLUSIONS

The binary ESN model herein introduced is in principle similar to regular ESNs. However, its simplicity permits a theoretical analysis of some important aspects of the transition to chaos. The expression we derived here for the autonomous case perfectly matches the experimental results. Our analysis of the noise applied to neuron activations showed how the network stability increases linearly with the mean degree of recurrent connections. The effects of input signals on the network dynamics are more complex to understand, since they introduce correlations among neurons. Our analysis partially explained the role that the signal magnitude and the mean degree play in shaping the EoC of the non-autonomous case.

References

- [1] C. Baldassi et al. “Learning may need only a few bits of synaptic precision”. In: *Physical Review E* 93.5 (2016), p. 052313.
- [2] Y. Cheng et al. “Model Compression and Acceleration for Deep Neural Networks: The Principles, Progress, and Challenges”. In: *IEEE Signal Processing Magazine* 35.1 (2018), pp. 126–136.
- [3] B. Derrida and Y. Pomeau. “Random networks of automata: a simple annealed approximation”. In: *EPL (Europhysics Letters)* 1.2 (1986), p. 45.
- [4] F. Farkhooi and W. Stannat. “Complete Mean-Field Theory for Dynamics of Binary Recurrent Networks”. In: *Physical review letters* 119.20 (2017), p. 208301.
- [5] C. Gallicchio and A. Micheli. “Echo State Property of Deep Reservoir Computing Networks”. In: *Cognitive Computation* (2017), pp. 1–14.
- [6] H. Jaeger. “The “echo state” approach to analysing and training recurrent neural networks-with an erratum note”. In: *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report* 148.34 (2001), p. 13.
- [7] S. A. Kauffman. “Metabolic stability and epigenesis in randomly constructed genetic nets”. In: *Journal of theoretical biology* 22.3 (1969), pp. 437–467.
- [8] C. G. Langton. “Computation at the edge of chaos: phase transitions and emergent computation”. In: *Physica D: Nonlinear Phenomena* 42.1-3 (1990), pp. 12–37.
- [9] R. Legenstein and W. Maass. “Edge of chaos and prediction of computational performance for neural circuit models”. In: *Neural Networks* 20.3 (2007), pp. 323–334.
- [10] L. Livi, F. M. Bianchi, and C. Alippi. “Determination of the edge of criticality in echo state networks through Fisher information maximization”. In: *IEEE Transactions on Neural Networks and Learning Systems* 29.3 (Mar. 2018), pp. 706–717.
- [11] K. Rajan, L. F. Abbott, and H. Sompolinsky. “Stimulus-dependent suppression of chaos in recurrent neural networks”. In: *Physical Review E* 82.1 (2010), p. 011903.
- [12] A. Rivkind and O. Barak. “Local dynamics in trained recurrent neural networks”. In: *Physical Review Letters* 118.25 (2017), p. 258101.
- [13] T. Rohlf and S. Bornholdt. “Criticality in random threshold networks: annealed approximation and beyond”. In: *Physica A: Statistical Mechanics and its Applications* 310.1-2 (2002), pp. 245–259.
- [14] X. R. Wang, J. T. Lizier, and M. Prokopenko. “Fisher information at the edge of chaos in random Boolean networks”. In: *Artificial Life* 17.4 (2011), pp. 315–329.
- [15] I. B. Yildiz, H. Jaeger, and S. J. Kiebel. “Re-visiting the echo state property”. In: *Neural Networks* 35 (2012), pp. 1–9.