

# From “Dynamics on Graphs” to “Dynamics of Graphs”: An Adaptive Echo-State Network Solution (Student Abstract)

Lei Zhang,<sup>1</sup> Zhqian Chen,<sup>2</sup> Chang-Tien Lu,<sup>1</sup> Liang Zhao<sup>3</sup>

<sup>1</sup> Virginia Tech, 7054 Haycock Rd, Falls Church, VA 22043, USA

<sup>2</sup> Mississippi State University, USA

<sup>3</sup> Emory University, USA

zhanglei@vt.edu (571-355-2886), zchen@cse.msstate.edu, ctlu@vt.edu, liang.zhao@emory.edu

## Abstract

Many real-world networks evolve over time, which results in dynamic graphs such as human mobility networks and brain networks. Usually, the “dynamics *on* graphs” (e.g., node attribute values evolving) are observable, and may be related to and indicative of the underlying “dynamics *of* graphs” (e.g., evolving of the graph topology). Traditional RNN-based methods are not adaptive or scalable for learning the unknown mappings between two types of dynamic graph data. This study presents a AD-ESN, and adaptive echo state network that can automatically learn the best neural network architecture for certain data while keeping the efficiency advantage of echo state networks. We show that AD-ESN can successfully discover the underlying pre-defined mapping function and unknown nonlinear map-ping between time series and graphs.

## Introduction

Graphs are widely utilized as universal representations of many real-world entities, such as social networks, brain functional connectivity, and molecule topology. These real-world networks typically involve patterns to their dynamics, which can usually be categorized as “dynamics *on* graphs” and “dynamics *of* graphs”. The former focuses on the time-evolving patterns of the entities’ activity, which can be demonstrated directly as the observed node attributes, while the latter emphasizes the underlying change in the topological structure of the graph. Existing methods are all designed under strong priors for specific assumptions for the graph definition, for example, interaction (Tupikina et al. 2016), entropy (Hlinka et al. 2013), and causal(Kipf et al. 2018).

This study aims to solve the following significant technical challenges: 1) Difficult in jointly considering the correlation of node and edge dynamics and their transformations. 2) Lack of efficient and scalable framework for graph dynamics encoding in large time duration with high resolution.3) Dilemma between model learnability and efficiency.

To address the above challenges, we present a novel framework to learn the mapping from dynamics on graphs to dynamics of graphs, by proposing a dynamic graph transformation model driven by a new adaptive deep echo-state

architecture. Specifically, a new deep architecture of echo-state network is proposed to efficiently encode the long time series of node attributes into dynamic edge embeddings. To further cope with the large search space of neural architectures of deep echo-state network, we formalize the deep echo-state neural architecture and propose new architecture search techniques.

## Methods

The base model of AD-ESN is the echo state network (ESN) (Lukoševičius and Jaeger 2009) based encoder which can be considered as a recurrent neural network where all of the weights are randomized and untrained. ESN does not *learn* the representation of the input time series. It is directly applied to the input and maps the input into a high-dimensional space. Different from vanilla ESNs, the input and internal node wiring patterns, where the weights are not zero, are formulated as the search space of deep ESN architectures and are optimized with neural architecture search (NAS). ESN and NAS together enable us to learn meaningful embeddings of arbitrary node signals in a graph.

After the “dynamics on graphs” are transformed into hidden representations with ESN, we apply an attention-based dynamic graph topological decoder that transforms the time series embeddings to dynamic graph topologies.

In the NAS stage, we decouple the deep ESN-based encoder and the graph topological decoder, then define the a surrogate loss function for the prediction performance of the ESN-based encoder. The architecture of the ESN-based encoder is optimized through the surrogate loss by using Gumbel-Softmax and gradient descent. The the training stage, the weights of the deep ESN-based encoder are randomized only on the edges determined in the NAS stage. The output weights in the deep ESN-based encoder and all the weight in the attention-based dynamic graph topological decoder are optimized as normal. Technical details are described in supplementary file.

## Experiments

### Datasets

We evaluate the effectiveness of the proposed model on two synthetic datasets. The synthetic chaotic time series data are generated with ground-truth signal to structure mappings.

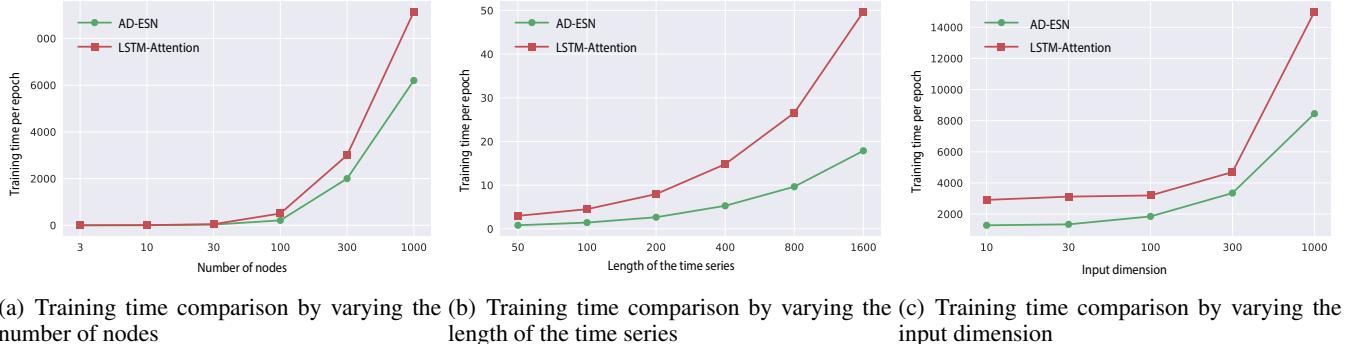


Figure 1: Scalability Analysis. The proposed AD-ESN achieves better scalability than the comparison method LSTM-Attention.

The synthetic weakly coupled time series data are generated with ground-truth structure to signal mappings. The two datasets evaluate the proposed model’s performance for reasoning and reverse-reasoning from “dynamics on graphs” to “dynamics of graphs”.

## Experiment Results

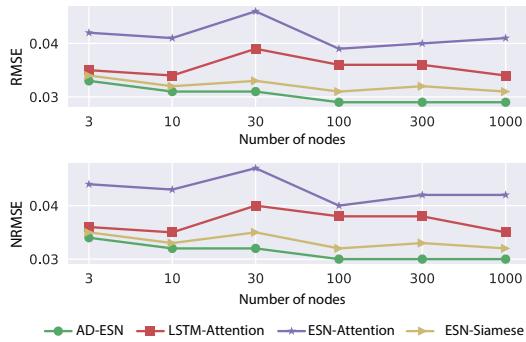


Figure 2: Performance comparison on Syn-Chaotic data.

As the Syn-Chaotic data is the easiest to generate, we compared the performance of AD-ESN along with three baseline models on this dataset, with a number of nodes ranging from 3 to 1,000. For the baseline methods, we replace the deep ESN encoder module in AD-ESN with LSTM (LSTM-Attention) and vanilla ESN (ESN-Attention), and replace the attention module with Siamese network. As is shown in Fig. 2, the proposed AD-ESN model achieves the best performance.

On the Syn-Coupled dataset, as shown in table 1, LSTM-

Method	RMSE	NRMSE	MAE	NMAE
LSTM-Attention	0.058	0.146	<b>0.050</b>	<b>0.125</b>
ESN-Attention	0.060	0.151	0.051	0.129
ESN-Siamese	0.602	1.505	0.599	1.498
AD-ESN	<b>0.057</b>	<b>0.144</b>	<b>0.050</b>	<b>0.125</b>

Table 1: Results on Syn-Coupled Data

Attention and AD-ESN got the same results. It means that the optimized ESN-based encoder can achieve the same performance as LSTM but with reduced computation.

Speed and high scalability are two significant advantages of ESNs. Fig. 1 presents a training time comparison between AD-ESN and LSTM-Attention. It is clear to see that the training time of AD-ESN increases much slower than LSTM-Attention when the scale of data increases. It is worth noting that the NAS stage of AD-ESN also takes time to train but it’s considered as a “once and for all” hyper parameter tuning process that runs automatically. Besides the training time, AD-ESN is also much more scalable in terms of GPU memory usage. We have a complexity analysis for both time and storage in the supplementary file.

## Conclusion and Future Work

We demonstrated that AD-ESN is capable of learning various of “dynamics on graphs” to “dynamics of graphs” mappings with minimum domain knowledge. We argue that ESNs can be powerful modules when they are pre-optimized with NAS while keeping the great scalability advantage. Future work comprises conducting more comparisons on both synthetic data and real-world data.

## References

- Hlinka, J.; Hartman, D.; Vejmelka, M.; Runge, J.; Marwan, N.; Kurths, J.; and Paluš, M. 2013. Reliability of inference of directed climate networks using conditional mutual information. *Entropy*, 15(6): 2023–2045.
- Kipf, T.; Fetaya, E.; Wang, K.-C.; Welling, M.; and Zemel, R. 2018. Neural relational inference for interacting systems. In *International Conference on Machine Learning*, 2688–2697. PMLR.
- Lukoševičius, M.; and Jaeger, H. 2009. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3): 127–149.
- Tupikina, L.; Molkenthin, N.; López, C.; Hernández-García, E.; Marwan, N.; and Kurths, J. 2016. Correlation networks from flows. The case of forced and time-dependent advection-diffusion dynamics. *PloS One*, 11(4): e0153703.