Deep connectomics networks: Results from neural network architectures inspired from network neuroscience

Anonymous Authors[1]

Abstract

Claims from the fields of network neuroscience and connectomics suggest that topological models of the brain involving complex networks are of particular use and interest. The field of deep neural networks has mostly left inspiration from these claims out. In this paper, we propose three architectures and use each of them to explore the intersection of network neuroscience and deep learning in an attempt to bridge the gap between the two fields. Using the teachings from network neuroscience and connectomics, we show improvements over the ResNet architecture, we show a possible connection between early training and the spectral properties of the network, and we show the trainability of a DNN based on the neuronal network of C.Elegans.

1. Introduction

1.1. Phenomena observed:

1. Improved classification accuracy obtained when we used deep neural network architectures whose topologies were inspired by network neuroscience. *Specifics*: 89.41% top-1 accuracy on CIFAR-10 with `Small-world RogueNet`

2. Positive correlation between the spectral gap of the graph and the first-epoch accuracy obtained during the training phase.
   Specifics: Pearson correlation of 0.65 on ImageNette with *Ramanujan-Net* topology

3. Freezing all the convolutional layers and training only the parameters associated with the graph edges still yield far better accuracy than random guess. *Specifics:* Training only the edge-weights of the C.Elagans neuronal network still resulted in 55.7% top-1 accuracy on the MNIST dataset.

4. Retaining of the small-world topology throughout the training phase.
   Specifics: Small-worldness of the topology measured by the small-world propensity metric used for weighted graphs was > 0.53 for all epochs of the training period for the MNIST dataset and the C.Elgenans neuronal network.

1.2. Background

There exists a formidable body of work in network neuroscience (Bassett & Sporns, 2017) and connectomics (Fornito et al., 2016) that deals with topological modeling of the brain's anatomical and functional networks. Some of the main claims from this field are:

1. Complex networks with small world topology (Watts & Strogatz, 1998) serve as an *attractive model for the organization of brain anatomical and functional networks because a small-world topology can support both segregated/specialized and distributed/integrated information processing* (Bassett & Bullmore, 2006)

2. *Human brain structural and functional networks follow small-world configuration and this small-world model captures individual cognition and exhibits physiological basis.* (Liao et al., 2017)

3. *Small-world, modules and hubs are present during the mid-gestation period. Early brain network topology can predict later behavioral and cognitive performance.*(Zhao et al., 2018; Watts & Strogatz, 1998)

It is therefore somewhat striking that the field of deep neural networks, with all its neuro-biologically inspired building blocks [1], has mostly left the *topology story* out.
There have been some nascent attempts worth a mention

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

[1] These sentences appear verbatim in (LeCun et al., 2015): *The convolutional and pooling layers in ConvNets are directly inspired by the classic notions of simple cells and complex cells in visual neuroscience, and the overall architecture is reminiscent of the*

here. In (Simard et al., 2005) the authors investigated multi-layered feed-forward networks trained with back propagation and found that the networks with the small-world connectivity resulted in lower learning error and lower training time when benchmarked with networks of regular or random connectivity. We note in passing that some specifics of the experimentation used in this paper have been challenged in (Guo, 2010). Continuing with small-world feed-forward neural networks theme, the authors in (Erkaymaz et al., 2017) obtained better results for a diabetes diagnosis dataset when compared with plain-vanilla fully connected feed-forward neural networks.

But in the post-AlexNet era, ever with this seemingly cambrian- explosion of neural network architectures [2], none seem to be inspired by the ideas prevalent in the domain of brain connectomics.

In this paper, we showcase a series of experimental attempts we've made to bridge the gap between the two communities (namely connectomics/network-neuroscience and deep learning). While it is not yet clear as to what deep neural network module would serve to be proverbial *node* in the complex network framework (we initially tried ResNet modules - See `RogueNet` section below), we firmly believe that the recent work by (Xie et al., 2019) serves as a truly elegant template and is worthy as the de facto choice for further modeling.

The rest of the paper is organized as follows: In section 2, we present RogueNet. A DNN model architecture with small-world topological properties where we used a ResNet module to be a node. In section-3, we present two architectures, RamanujanNet and C.ElgansNet based on the node model proposed in (Xie et al., 2019). In section-4, we conclude the paper and present the current directions of research we are currently pursuing to extend this work.

2. Node model-1: RogueNet

Our proposed RogueNet architecture expands upon the Pre-Activation ResNet architecture in which ReLu activations and batch normalization are performed before convolutions (He et al., 2016). We introduce *random additive connections* between ResNet modules in a global manner according to a connectivity pattern resembling a Watts-Strogatz small world network. The additive skip connections are combined with the activation from a given ResNet module using a convolutional gating mechanism inspired by Highway Networks (Srivastava et al., 2015).

LGN-V1-V2-V4-IT hierarchy in the visual cortex ventral pathway. When ConvNet models and monkeys are shown the same picture, the activations of high-level units in the ConvNet explains half of the variance of random sets of 160 neurons in the monkeys inferotemporal cortex.

[2] See `http://www.asimovinstitute.org/neural-network-zoo/` to visualize the growing neural-network zoo

Table 1. Types of ResNet modules used in RogueNet

Basic	Bottleneck
Batch Normalization	Batch Normalization
ReLU	ReLU
Convolution (3x3)	Convolution (1x1)
Batch Normalization	Batch Normalization
ReLU	ReLU
Convolution (3x3)	Convolution (3x3)
	Batch Normalization
	ReLU
	Convolution (1x1)

2.1. Topology of the RogueNet architecture

The base architecture which we apply random connections to is composed of Pre-Activation ResNet modules. These modules come in two flavors as proposed by He et al., the first of which is referred to as a 'basic' module which contains two convolutional layers, and the second of which is referred to as a 'bottleneck' module which contains three convolutional layers (He et al., 2015). The details of each of these are presented in Table 1.

The experiments presented here rely primarily on the use of basic ResNet modules, and we compare RogueNet architectures consisting of 34 layers (excluding gating layers) to ResNet34, which similarly has 34 layers. Each ResNet module takes input from the previous ResNet module or layer and skip connections from previous layers in the network pass activations which are added to it before being passed to a gating mechanism and being sent to one or more later layers in the network. The connectivity pattern of the skip connections is a Watts Strogatz random network with mean edge density and rewiring probability set such that it has small world properties. Our neuron architecture can be seen in Figure 3

The gating mechanism used in the RogueNet architecture is inspired by Highway layers, introduced by Srivastava et al. (Srivastava et al., 2015). The original formulation for Highway layers is expressed as

$$y = H(x, W_H) \cdot T(x, W_T) + x \cdot (1 - T(x, W_T))$$

where $H(x, W_H)$ is an arbitrary fully connected layer with input x and parameters W_H. $T(x, W_T)$ is specified to be a fully connected layer parameterized by W_T and a bias vector with input x with the sigmoid activation function.

Our gating mechanism differs from this in that we use convolutional layers, and our transformation $H(x, \cdot)$ sums the

output and the activations from the incoming skip connections. Mathematically, this is described as follows

$$y_L = \left(x_L + \sum_{i \in S} x_i\right) \cdot T_{conv}(x_L, W_T) + x_L \cdot (1 - T_{conv}(x_L, W_T)),$$

where L is the index of the current layer and S is the set of indices of layers which have a skip connection to layer L. X_j where j refers to the index of a particular layer, is the activation from the gating mechanism from layer j. In this case, $T_{conv}(x_L, \cdot)$ is a 1 by 1 convolution with a bias term and the sigmoid activation function. Intuitively, this layer learns how much the incoming skip connection activations should be weighted relative to the activation from the previous module.

2.2. Training procedure for RogueNet

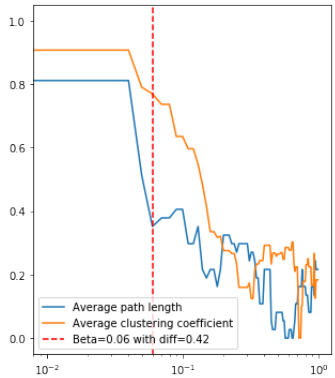

Figure 1. Estimating β for a 17 node small world network with $K = 6$.

We use the same training procedure for each of the RogueNet experiments, including our comparison to ResNet. We trained for 200 epochs on the CIFAR-10 dataset using SGD with a learning rate of 0.1, momentum set to 0.9, weight decay set to 0.0005, and a batch size of 128. We randomly re-compute the connectivity pattern of the skip connections once per epoch according to the Watts Strogatz recipe for generating random Watts Strogatz graphs. This process is parameterized by a mean edge density, K, and a rewiring probability, β, which were fixed to be $K = 4$ and $\beta = 0.08$. K was found empirically, and β was estimated based on the value of K by generating random graphs with various values of β and choosing the β yielding the greatest difference in the normalized average clustering coefficient and the average path length, which are properties of small world networks. An example result of this computation can be seen in Figure 1

Table 2. Performance comparison on CIFAR-10.

Architecture ($K = 4$ and $\beta = 0.08$ for RogueNet34)	Test accuracy
ResNet34	88.000%
RogueNet34 (highway layer, no shuffling)	88.510%
RogueNet34 (frozen highway layer, no shuffling)	88.402%
RogueNet34 (pre-activations, highway layer, no shuffling)	88.610%
RogueNet34 (pre-activations, highway layer, epoch-wise shuffling)	89.410%

2.3. RogueNet results

In our RogueNet experiments, we explored various configurations of using pre-activations, freezing the highway layer, and shuffling the random skip connections. We compare our top performing configurations with ResNet34 using the same training procedure on CIFAR-10. We find that our best configuration involves using pre-activations, using trainable parameters in the gating mechanisms, and shuffling the skip connections once per epoch. This and the results of the rest of the configurations can be found in Table 2. This configuration achieves a test accuracy on CIFAR-10 of 89.410%, where ResNet34 achieves a test accuracy of 88.000%.

3. Node model-2: RamanujanNet and C.EleganSNet

3.1. RamanujanNet

We propose an architecture which leverages the spectral properties of the connectivity pattern in the graph, based on the RandWire architecture proposed by Xie et al. (Xie et al., 2019). In their work, they propose an architecture composed of three randomly connected graph modules, each with random connections between nodes consisting of a single convolutional layer. In their work, they propose the use of Watts Strogatz, Erdos Renyi, and Barabasi Albert graphs, where a different random graph is used for each of the three modules. We propose the use of expander graphs where the same graph is replicated across the three modules. We experiment with expander graphs with various spectral gaps generated by sampling random K-regular graphs and computing the spectral gap. In our experiments, each graph consists of 36 nodes and 64 edges.

3.2. Training procedure for RamaujanNet

We train the RamaujanNet architecture using 9 different graph topologies with various spectral gaps for 1 epoch each on MNIST and ImageNette using the Adam optimizer

with a learning rate of 0.001 (Kingma & Ba, 2014).

3.3. RamaujanNet results

We evaluated the first epoch accuracy across randomly sampled K-regular graphs with various spectral gaps. We discovered that the first epoch accuracy was positively correlated with spectral gap across the MNIST and Imagenette datasets (fas). The results for these can be found in Figure 2. Here, we show that the Pearson correlation between first epoch accuracy and spectral gap on MNIST is 0.55, and the Pearson correlation on the Imagenette dataset is 0.65.

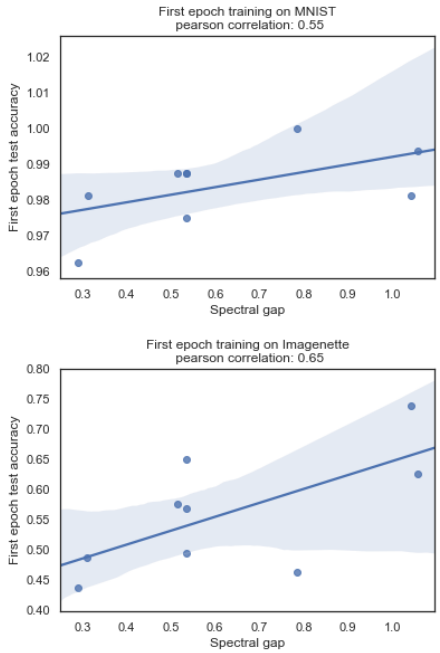

Figure 2. First epoch accuracy by spectral gap on MNIST and Imagenette.

3.4. C.EleganSNet

We propose another architecture which is also based on RandWire, but this time using only a single graph based module instead of three such consecutive modules as described by Xie et al. (Xie et al., 2019). In this architecture, we choose our participant graph topology to be a small world network found in nature, namely the neuronal network of C.Elegans. In this series of experiments, we trained a neural network using this architecture on MNIST, KMNIST, and Fashion MNIST. An additional experiment was conducted in which we froze every parameter including the final fully connected softmax layer except those corresponding to edge weights in the C.Elegans small world network, and tracked the evolution of small world propensity of the graph over training.

3.5. Training procedure for C.EleganSNet

We train the C.EleganSNet architecture with 10 random initializations for 2 epochs each on MNIST, KMNIST, and Fashion MNIST using the Adam optimizer with a learning rate of 0.001 (Kingma & Ba, 2014). For our experiment in which we froze all other parameters except for the network edge weights, we trained on MNIST for 20 epochs. Additionally, we tracked the small world propensity of the C.Elegans graph with the learned edge weights over training. Before computing the small world propensity, we standardized the neural network parameters and applied the sigmoid activation function to ensure positivity while transforming the parameters in a monotonic way.

3.6. C.EleganSNet results

We discovered that the C.EleganSNet architecture was able to consistently achieve competitive results on MNIST, KMNIST, and Fashion MNIST. Figure 4 shows the distribution of our results compiled from 10 training trials on each dataset. The mean test accuracy on MNIST was 99%, while we achieved 93% on KMNIST, and 90% on Fashion MNIST.

Additionally, upon freezing all parameters traditionally associated with deep neural networks leaving the edge weights as the only trainable parameters, we found that the C.EleganSNet architecture achieved 55.7% accuracy on MNIST. The loss and accuracy curves for this experiment can be seen in Figure 5. For this experiment we also tracked the small world propensity of the C.Elegans graph with the learned edge weights over training and found that it does not significantly change throughout training.

4. Conclusion and future work

We have demonstrated three distinct approaches to applying work from network neurosciences and connectomics to deep learning. Our experiments show improvements over ResNet by the inclusion of skip connections which follow a connectivity pattern with small world properties, a possible connection between early training performance and spectral gap when using expander graphs as the participant graph topology with the node model proposed by (Xie et al., 2019), and the trainability of a DNN based on the neuronal network of C.Elegans with and without freezing the parameters of the convolutional and fully connected layers.

In future work, we will examine the impact of other spectral properties of the graph topologies used both in the architectures we proposed and in the RandWire architecture proposed by (Xie et al., 2019). Additionally, we will explore parameter efficient connectivity patterns which could achieve similar performance to related networks with more parameters.

References

fastai/imagenette: A smaller subset of 10 easily classified classes from imagenet, and a little more french. https://github.com/fastai/imagenette. (Accessed on 05/24/2019).

Bassett, D. S. and Bullmore, E. Small-world brain networks. *The neuroscientist*, 12(6):512–523, 2006.

Bassett, D. S. and Sporns, O. Network neuroscience. *Nature neuroscience*, 20(3):353, 2017.

Erkaymaz, O., Ozer, M., and Perc, M. Performance of small-world feedforward neural networks for the diagnosis of diabetes. *Appl. Math. Comput.*, 311(C):22–28, October 2017. ISSN 0096-3003. doi: 10.1016/j.amc.2017.05.010. URL https://doi.org/10.1016/j.amc.2017.05.010.

Fornito, A., Zalesky, A., and Bullmore, E. *Fundamentals of brain network analysis*. Academic Press, 2016.

Guo, Z. Comment on" fastest learning in small-world neural networks". *arXiv preprint arXiv:1003.0060*, 2010.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015. URL http://arxiv.org/abs/1512.03385.

He, K., Zhang, X., Ren, S., and Sun, J. Identity mappings in deep residual networks. *CoRR*, abs/1603.05027, 2016. URL http://arxiv.org/abs/1603.05027.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization, 2014.

LeCun, Y., Bengio, Y., and Hinton, G. Deep learning. *Nature*, 521(7553):436–444, May 2015. ISSN 0028-0836. doi: 10.1038/nature14539. URL http://dx.doi.org/10.1038/nature14539.

Liao, X., Vasilakos, A. V., and He, Y. Small-world human brain networks: perspectives and challenges. *Neuroscience & Biobehavioral Reviews*, 77:286–300, 2017.

Simard, D., Nadeau, L., and Kröger, H. Fastest learning in small-world neural networks. *Physics Letters A*, 336(1):8–15, 2005.

Srivastava, R. K., Greff, K., and Schmidhuber, J. Highway networks. *CoRR*, abs/1505.00387, 2015. URL http://arxiv.org/abs/1505.00387.

Watts, D. J. and Strogatz, S. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, June 1998.

Xie, S., Kirillov, A., Girshick, R., and He, K. Exploring randomly wired neural networks for image recognition, 2019.

Zhao, T., xu, Y., and He, Y. Graph theoretical modeling of baby brain networks. *NeuroImage*, 06 2018. doi: 10.1016/j.neuroimage.2018.06.038.

A. Additional figures

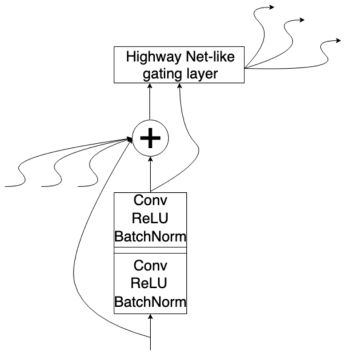

Figure 3. Neuron architecture for RogueNet.

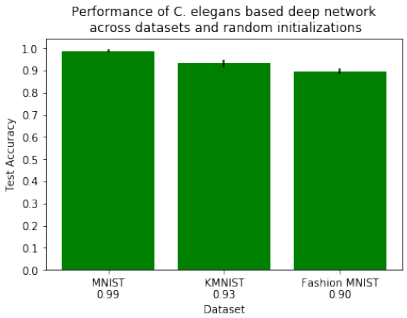

Figure 4. Distribution of accuracies of C.EleganSNet on MNIST, KMNIST, and Fashion MNIST.

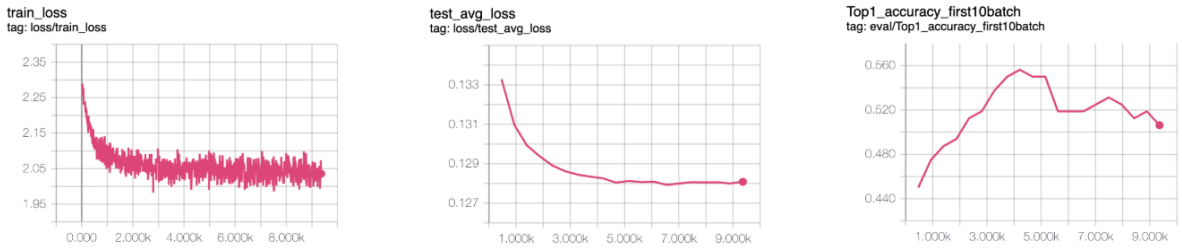

Figure 5. Performance of C.EleganSNet with all parameters frozen except the C.Elegans graph edge weights.

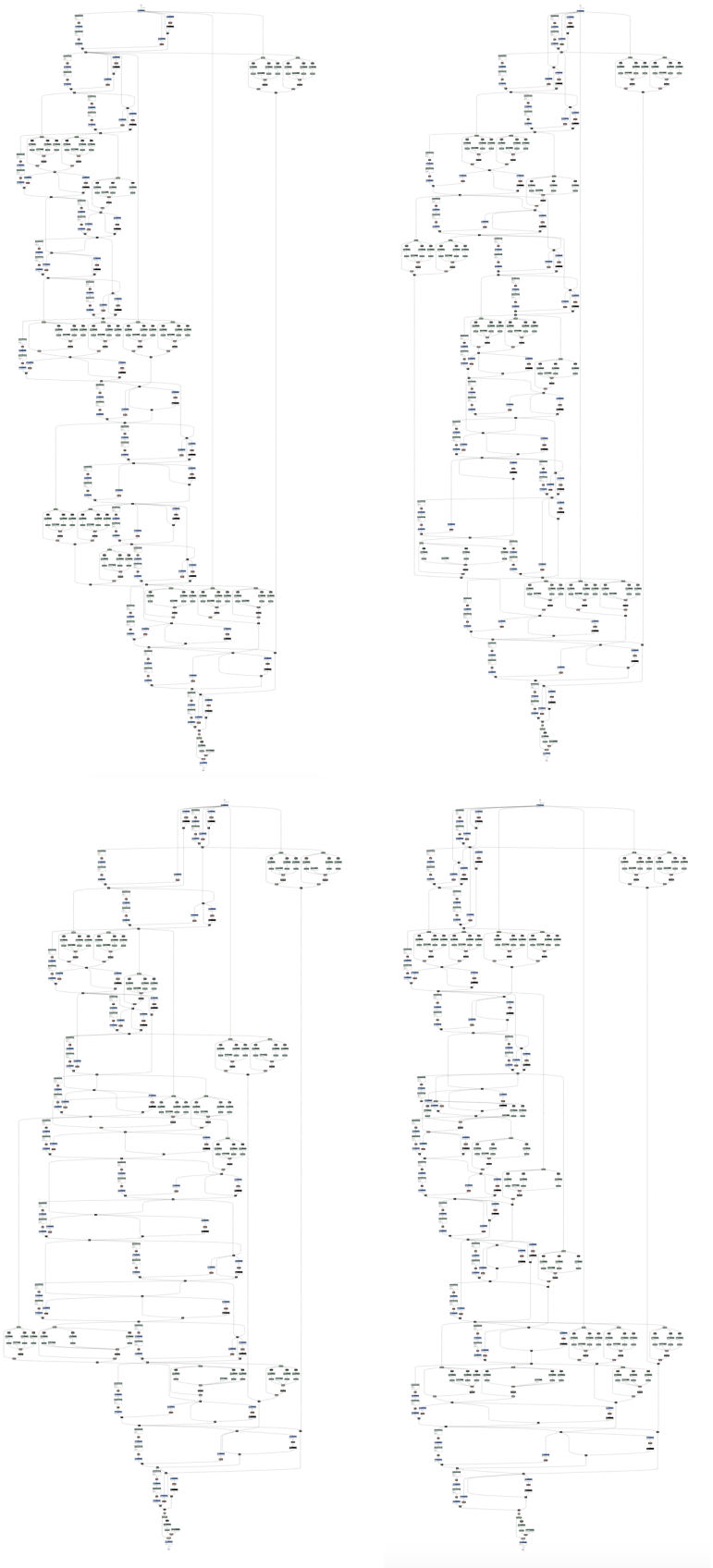

Figure 6. Examples of random small world skip connections using a base architecture of ResNet34.

Figure 7. High level architecture of C.EleganSNet.

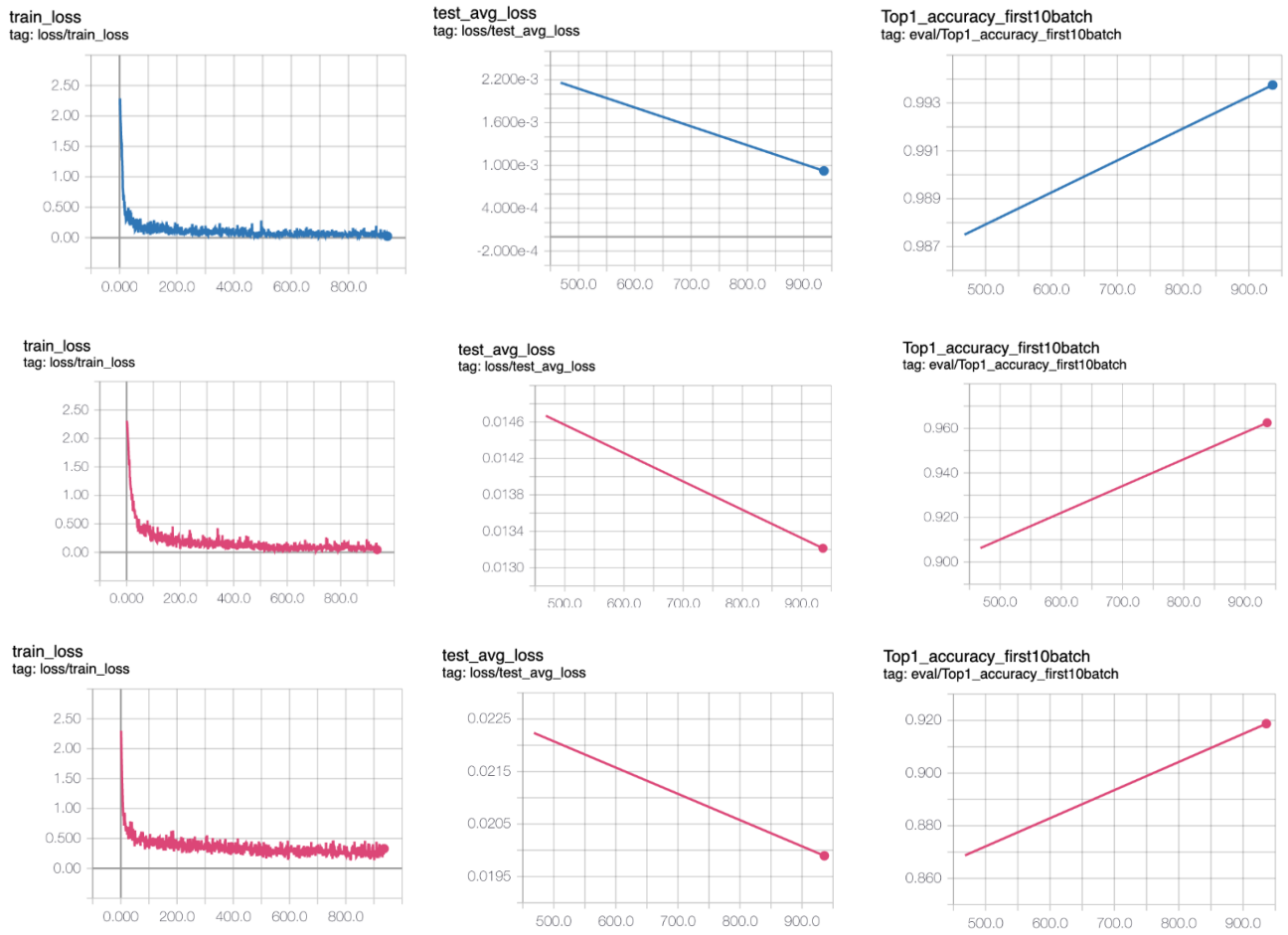

Figure 8. Loss and accuracy curves for C.EleganSNet with no parameter freezing (Top to bottom: MNIST, KMNIST, Fashion MNIST; Left to right: Training loss, test loss, test accuracy).

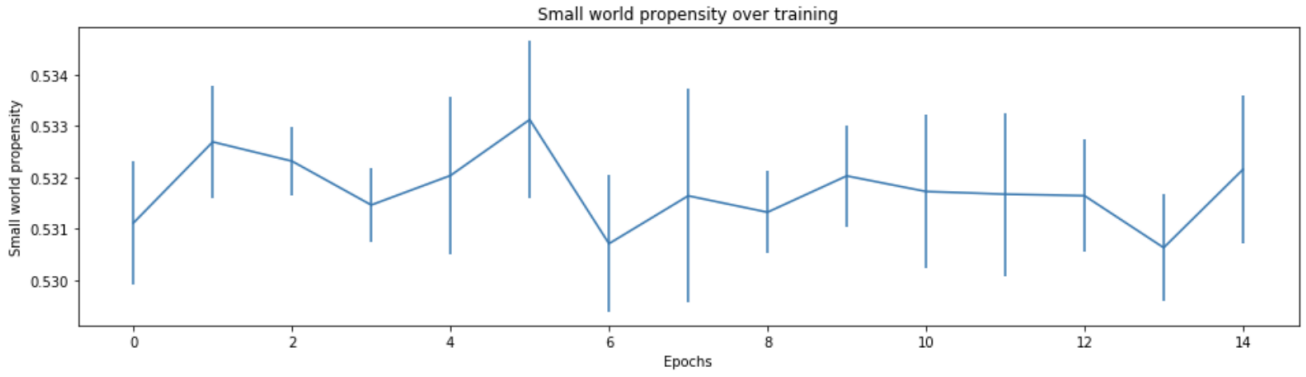

Figure 9. Evolution of the small world propensity of the C.Elegans graph throughout training on MNIST with frozen non-edge weight parameters.