Robustness in sparse artificial neural networks trained with adaptive topology

Bendegúz Sulyok[1, 2], Gergely Palla[1, 2] Filippo Radicchi[3], and Santo Fortunato[3]

[1]Dept. of Biological Physics, Eötvös University, Budapest, Hungary

[2]Health Services Management Training Centre, Semmelweis University, Budapest, Hungary

[3]Center for Complex Networks and Systems Research Luddy School of Informatics, Computing, and Engineering Indiana University Bloomington, USA

Modern deep learning faces significant challenges regarding energy consumption and computational costs due to the massive over-parameterization of standard architectures. This research explores a biologically inspired alternative: highly sparse artificial neural networks where only 1% of possible connections are maintained, mimicking the vanishingly small connectivity fraction observed in the human brain. By utilizing an "epitopological" learning approach, the network's topology is dynamically updated after each training epoch. This allows the model to "prune" unimportant connections and "regrow" new ones, effectively searching for an optimal architecture during the training process rather than relying on a fixed, dense structure.

The study compares two primary regrowth strategies—Random Link Regrowth (RLR) and the CH3L3 heuristic—across four image classification benchmarks: MNIST, Fashion MNIST, KMNIST, and EMNIST Letters. The CH3L3 method is based on link prediction and evaluates the likelihood of missing links to guide connectivity. While both methods achieve competitive final accuracy despite the 99% reduction in parameters, the CH3L3 method demonstrates significantly faster convergence during training. The central contribution of this work, however, is a detailed analysis showing that these adaptive sparse networks are remarkably robust to post-training disruptions.

To assess resilience, the networks were subjected to iterative structural pruning (random and weight-ordered) as well as stochastic weight perturbations such as shuffling and noise addition. Under "Reverse Weight Order Pruning," where links are removed starting from the smallest weights, the networks exhibit a graceful degradation of performance. A key highlight is the "structural quality" inherent in the RLR-trained networks; these models retain high accuracy even when 80% of their remaining links are removed, suggesting that the random training process embeds information in a highly distributed and resilient manner.

Ultimately, this work demonstrates that the specific dynamic training method employed has a notable effect on the resulting network's structural properties and resilience. While CH3L3 offers speed, RLR yields models that are inherently more stable and compressible. These findings highlight the potential of adaptive sparse networks as a path toward developing efficient, sustainable, and reliable deep learning models for resource-constrained environments.

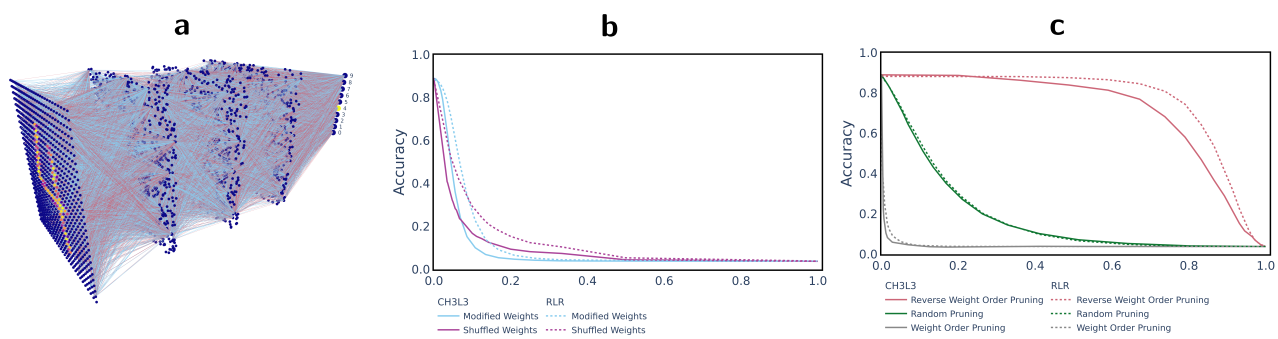

Figure 1: **(a) Network Architecture.** Illustration of the 99% sparse network architecture. The input layer (shown on the left) contains 28 x 28 pixels, followed by 3 sparsely connected neuron layers with 1000 neurons each. The last layer, providing the readout is densely connected to the third sparse layer. **(b-c) Robustness results.** The results of applying various perturbation strategies to the trained networks demonstrate that while the networks are highly sensitive to the removal of high-magnitude weights, they exhibit significant resilience under random and reverse-weight perturbations.

[1] Carlo Vittorio Cannistraci. Modelling self-organization in complex networks via a brain-inspired network automata theory improves link reliability in protein interactomes. Scientific Reports, 8(1), oct 2018.

[2] Frankle, J. & Carbin, M. (2019). The lottery ticket hypothesis: Finding sparse, trainable neural networks. International Conference on Learning Representations (ICLR).