A Dendritic-Inspired Network Science Generative Model for Topological Initialization of Connectivity in Sparse Artificial Neural Networks

Diego Cerretti[1,2], Yingtao Zhang[1,2], Carlo Vittorio Cannistraci[1,2,3]

[1]Center for Complex Network Intelligence (CCNI)

[2]Department of Computer Science, [3]Department of Biomedical Engineering, Tsinghua University

diegocerretti02@gmail.com, zhangyingtao1024@gmail.com, kalokagathos.agon@gmail.com

清華大學 脑与智能实验室

Tsinghua Laboratory of Brain and Intelligence

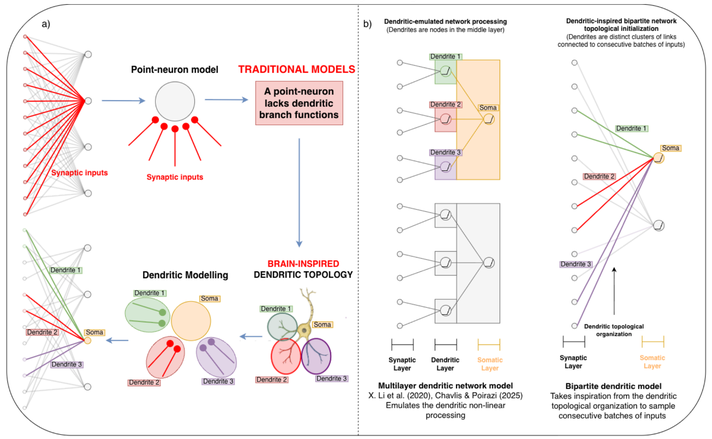
NEURAL INFORMATION PROCESSING SYSTEMS

Abstract

We introduce the Dendritic Network Model (DNM), a generative framework for creating ultra-sparse, bio-inspired networks. Instead of random initialization, DNM uses parametric distributions, optimized via network science, to define topology. DNM consistently outperforms standard methods at 99% sparsity in image classification and improves sparse Transformers for machine translation, offering a structural advantage for more scalable and efficient AI.

The Dendritic Network Model (DNM)

Inspired by biological neurons, the Dendritic Network Model (DNM) creates a structured connectivity pattern where each output neuron connects to the input layer via several distinct **"dendritic branches."** Each branch links the output neuron to a **consecutive block** of input neurons. All of an output neuron's branches must form within a predefined **"receptive window"** of the input layer. This method results in **structured, clustered sparsity**.

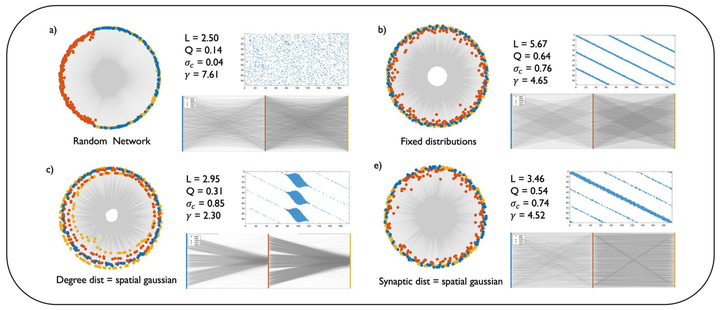


The Dendritic Network Model. Traditional models treat neurons as simple integrators. In contrast, our brain-inspired approach organizes connections into dendritic branches, creating a structured topology where distinct groups of inputs are processed locally. Compared to previous multi-layered dendritic models[4,5], DNM embeds dendritic properties directly into the bipartite network topology.

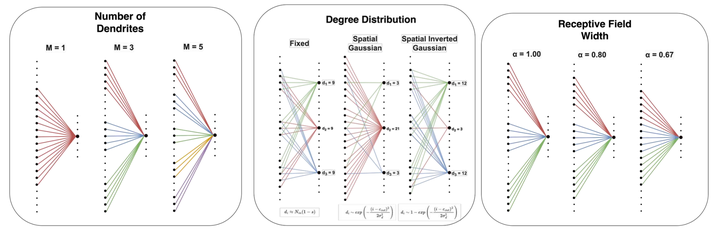
Experiments

We evaluate DNM on static and dynamic sparse training (DST). For static training, we test MLPs on image classification tasks (MNIST, Fashion MNIST, EMNIST, CIFAR-10) and benchmark DNM against random, BSW, BRF, Ramanujan, CSTI, and SNIP initializations. For DST, we compare against the same methods and use SET[1] and RigL[2] as baselines, while also testing DNM as an initializer for state-of-the-art CHTs and CHTss[3]. Finally, we evaluate DNM as an initializer for Transformers on machine translation tasks (Multi30k, IWSLT14, WMT17), comparing it with the BRF method.

	CSTI	Random	BSW	BRF	Ramanujan	DNM
FC	62.85±0.16					
CHTs	69.77±0.06	66.96±0.24	66.96±0.24	64.96±0.17	67.19±0.17	**68.76±0.11**
CHTss	71.29±0.14	66.89±0.23	66.89±0.23	64.96±0.17	67.37±0.12	**68.50±0.21**



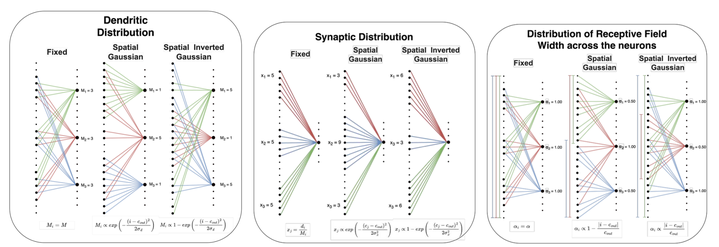
Geometric and topological characterization of the Dendritic Network Model. The figure compares a baseline random network (a) with various DNM configurations (b-d) for a 3-layered MLP of size 98x196x196 with 90% sparsity. Each panel shows a coalescent embedding in hyperbolic space (left), the first layer's adjacency matrix (top right), a bipartite graph representation (bottom right), and key network science metrics: characteristic path length (L), modularity (Q), structural consistency (σ_c), and the power law exponent of the degree distribution (γ). The network in (b) is a standard DNM model, generated using fixed distributions for all parameters. Panels (c-d) modify this standard configuration by switching a single parameter's distribution to spatial Gaussian: (c) degree distribution, (d) synaptic distribution.



Effect of the number of dendrites on the network topology

Effect of the degree distribution on the network topology

Effect of the receptive field width on the network topology



Effect of the dendritic distribution on the network topology

Effect of the synaptic distribution on the network topology

Effect of the receptive field distribution on the network topology

	MNIST		Fashion MNIST		EMNIST	
	CHTs	CHTss	CHTs	CHTss	CHTs	CHTss
FC	98.78±0.02		90.88±0.02		87.13±0.04	
CSTI	98.81±0.04	98.83±0.02	90.93±0.03	90.81±0.11	87.82±0.04	87.52±0.04
Random	98.57±0.04	98.61±0.03	90.42±0.03	90.30±0.10	87.12±0.13	87.20±0.09
BSW	98.57±0.04	98.61±0.04	90.46±0.06	90.46±0.06	87.12±0.13	87.20±0.09
BRF	98.47±0.03	98.47±0.03	90.04±0.12	90.04±0.12	87.03±0.07	87.03±0.07
Ramanujan	98.57±0.03	98.57±0.03	89.82±0.14	89.78±0.08	87.24±0.08	87.24±0.08
DNM	**98.66±0.03**	**98.90±0.01***	**90.68±0.09**	**90.88±0.03***	**87.40±0.04**	**87.55±0.01***

Image classification on MNIST, Fashion MNIST, and EMNIST of the CHTs and CHTss models on MLPs with 99% sparsity over various topological initialization methods, compared to the fully-connected (FC) model. The scores indicate the accuracy of the models, averaged over 3 seeds ± their standard errors. Bold values denote the best performance amongst initialization methods different from CSTI. The performances that surpass CSTI are marked with "*".

Model	WMT		Multi30k		IWSLT	
	0.95	0.90	0.95	0.90	0.95	0.90
FC	25.52		31.38±0.38		24.48±0.30	
CHTs[B]	20.94±0.63	22.40±0.06	28.94±0.57	29.81±0.37	21.15±0.10	21.92±0.17
CHTs[D]	**21.34±0.20**	**22.56±0.14**	**30.54±0.42**	**31.45±0.35***	**22.09±0.14**	**23.52±0.24**
CHTss[B]	23.73±0.43	**24.61±0.14**	32.03±0.29*	32.86±0.16*	**24.51±0.02***	**24.31±0.04**
CHTss[D]	**24.52±0.12**	24.40±0.15	**32.62±0.28***	**33.00±0.31***	24.43±0.14	24.20±0.07

[D]DNM initialization. [B]BRF initialization.

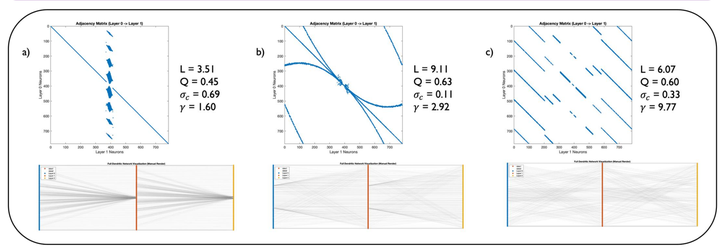
Performance comparison on machine translation tasks across the WMT, Multi30k, and IWSLT datasets For tests on WMT, the DNM model's parameters were transferred from the best-performing combinations of previous tests, avoiding any parameter search. Entries are BLEU scores (higher is better), averaged over 3 seeds ± standard error. Bold values denote the best performance for a given sparsity and initialization. Values that surpass the fully connected (FC) transformer are denoted by "*"

Static Sparse Training

	MNIST	Fashion MNIST	EMNIST	CIFAR10
FC	98.78±0.02	90.88±0.02	87.13±0.04	62.85±0.16
CSTI	98.07±0.02	88.52±0.14	84.66±0.13	52.64±0.30
SNIP	97.59±0.08	87.85±0.22	84.08±0.08	61.81±0.58
Random	96.72±0.04	87.34±0.11	82.66±0.08	55.28±0.09
BSW	97.32±0.02	88.18±0.18	82.94±0.06	56.54±0.15
BRF	96.85±0.01	87.41±0.13	82.98±0.02	54.73±0.07
Ramanujan	96.51±0.17	86.45±0.15	81.80±0.13	55.05±0.40
DNM	**97.82±0.03**	**89.19±0.01**	**84.76±0.13**	**61.63±0.18**

Image classification accuracy of statically trained, 99% sparse MLPs with different initial network topologies, compared to the fully-connected (FC) model. The scores are averaged over 3 seeds ± their standard errors. Bold values denote the best performance amongst initialization methods different from CSTI and SNIP.

Results Analysis



Representation of the best performing DNM models on image classification. The figure compares the best performing DNM architectures on MNIST and Fashion MNIST (a), EMNIST (b), and CIFAR10 (c). Each panel shows the network's adjacency matrix (top) and the network's layerwise representation (bottom). Furthermore, each panel exhibits the network's topological measures: characteristic path length (L), modularity (Q), structural consistency (σ_c), and the power law exponent of the degree distribution (γ).

Experiments show that **DNM consistently outperforms alternative topological initializers at extreme sparsity in both static and dynamic training**. Crucially, results suggest that simpler datasets favor hierarchical, scale-free structures, while complex visual data prefers distributed, non-hierarchical connectivity. This finding hints that the optimal sparse topology is task-dependent and establishes DNM as a principled platform for exploring the relationship between topology and function.

References

[1]Mocanu et al., Scalable training of artificial neural networks with adaptive sparse connectivity inspired by network science. Nature communications, 2028.

[2]Evci et al., Rigging the lottery: Making all tickets winners. PMLR, 2020.

[3]Y. Zhang et al., Brain network science modelling of sparse neural networks enables transformers and llms to perform as fully connected. arXiv, 2025.

[4]Chavlis, Spyridon, and Panayiota Poirazi. "Dendrites endow artificial neural networks with accurate, robust and parameter-efficient learning." Nature communications 16.1 (2025): 943.

[5]Li, Xinyi, et al. "Power-efficient neural network with artificial dendrites." Nature Nanotechnology 15.9 (2020): 776-782.