Paying the Price for Reach: Size-Dependent Emergence of Efficient Wiring in Cognitive Recurrent Neural Networks

Bilal Aslan[1], Geoff Nitschke[2]

aslbil001@myuct.ac.za[1] gnitschke@cs.uct.ac.za[2]
Department of Computer Science
University of Cape Town
Cape Town, South Africa

Abstract

Artificial neural networks often neglect the physical wiring costs that are crucial to biological nervous systems. This study investigates how incorporating such a biologically inspired wiring constraint influences the structure and function of Recurrent Neural Networks (RNNs) during learning. We systematically varied network size (N) and the strength (λ) of a communicability-weighted spatial regularization penalty applied to RNNs performing a seasonal foraging task requiring short and long-term memory, and decision-making. Our results reveal that while all networks achieved high task accuracy, larger networks ($N \geq 100$) exhibited different sensitivity patterns to higher penalties (λ) compared to smaller ones (N=50). Increasing λ induced neural network topologies with similarities to biological brains, including sparsity, shorter connection lengths (while preserving crucial long-range connections), increased modularity, and enhanced small-world characteristics. We identify a size-dependent optimum (*sweet spot*) for $\lambda \in [0.05, 0.10]$ that yields these efficient, brain-like structural properties without compromising functional performance. These results highlight the importance of physical network constraints in shaping adaptive systems, demonstrate how functional networks can self-organize towards efficient topologies under cost pressures and offer design principles for developing neuromorphic systems.

Introduction

Biological nervous systems represent paradigms of adaptive complexity, capable of sophisticated computation and learning within strict physical and metabolic budgets (Laughlin et al., 1998; Bullmore and Sporns, 2012). A fundamental constraint is *wiring cost*, the material and energetic expense of constructing and maintaining neuronal connections (axons and dendrites) within the brain's physical volume (Chklovskii et al., 2002). This pressure is thought to profoundly shape brain architecture, favoring efficient network topologies that balance computational needs with resource limitations (Striedter, 2004). Understanding how such constraints influence the emergence of structure and function is a core pursuit in both computational neuroscience and Artificial Life (ALife), which seeks to understand the essential principles governing living systems, including adaptation under various constraints (Gershenson, 2023).

Recurrent Neural Networks (RNNs) have been demonstrated as powerful tools for modeling cognitive processes over time, successfully capturing aspects of working memory, decision-making, and sequence processing (Barak, 2017; Greff et al., 2016; Yang et al., 2019). However, standard RNN models are typically abstract, neglecting the spatial embedding and associated wiring costs inherent to their biological counterparts. This omission limits their biological plausibility and overlooks fundamental organizational principles governing efficient computation in physical systems. While some studies have explored spatial factors in network models (Kaiser and Hilgetag, 2006; Achterberg et al., 2023), there is little work on how functionally-modulated wiring costs impact network scale during complex learning tasks.

This study addresses this gap by investigating the interplay between network size (N) and an explicit penalty mimicking wiring cost (λ) in RNNs. We implement a weighted spatial regularization term (Achterberg et al., 2023) that penalizes connection weights based on Euclidean distances between connected neurons, modulated by the connection's role in global network communication (Estrada and Hatano, 2008). We train RNNs of varying sizes ($N \in \{50, 100, 200\}$) on a seasonal foraging task designed to simultaneously engage working memory, long-term memory, and decision-making. This study's research objectives are:

1. Quantify the impact of network size (N) and spatial regularization strength (λ) on task performance.
2. Analyze and ascertain how λ influences the learned recurrent connectivity structure ($\boldsymbol{W}_{rec}$), given metrics for sparsity, connection length, modularity, and small-worldness.
3. Investigate the interaction between N and λ to understand how scaling affects the emergence of efficient neural network topologies given spatial constraints.

By systematically varying N and λ (Table 2) and analyzing the resultant network dynamics, structure, and task performance, we demonstrate that incorporating spatial costs

drives RNNs to develop topologies with key characteristics observed in biological brains, such as sparsity, modularity, and small-worldness. We find that the optimal level of spatial pressure is size-dependent, suggesting a crucial trade-off between network connection capacity, wiring economy, and task-accomplishing capacity. Results provide insights into the principles of constrained neural network adaptation and offer guidelines for designing more efficient and biologically plausible artificial cognitive architectures.

Background

Biological neural networks exhibit remarkable cognitive abilities given the coordinated activity of vast numbers of interconnected neurons (Kandel et al., 2000). However, unlike many abstract computational models, biological brains are physical entities subject to fundamental biophysical constraints, including spatial embedding and metabolic costs associated with building and maintaining neural connections (Laughlin and Sejnowski, 2003; Chklovskii et al., 2002).

RNNs for Sequential Processing and Cognition

RNNs process sequential data, making them powerful tools for modeling temporal dependencies (Elman, 1991; Hopfield, 1982). RNNs possess connections that form directed cycles, allowing them to maintain an internal state or *memory* to capture information from past inputs. This is crucial for modeling cognitive processes over time, such as language processing (Sutskever et al., 2014), evidence-based decision-making (Wang, 2002), and working memory (Greff et al., 2016). RNN architectures, such as the SimpleRNN (Elman, 1991), and more advanced variants like Long Short-Term Memory (LSTM) (Greff et al., 2016) and Gated Recurrent Units (GRU) (Cho et al., 2014), have been successfully simulated complex cognitive functions. For instance, RNNs trained on specific tasks can develop dynamics that mimic neural activity observed in biological experiments during working memory delays (Barak and Tsodyks, 2014) or context-dependent decision-making (Yang et al., 2019).

Spatial Embedding and Wiring Cost Constraints

While abstract RNN models capture temporal dynamics, they often neglect the physical reality of biological neural circuits. Neurons in the brain occupy physical space, and their connections (axons and dendrites) have associated material and metabolic costs that increase with length (Azevedo et al., 2009; Bullmore and Sporns, 2012). This *wiring cost* is considered a significant evolutionary pressure shaping brain architecture (Chklovskii et al., 2002; Striedter, 2004). Empirical evidence supports the idea that brain networks are organized to minimize wiring costs while maintaining efficient information processing, often resulting in modular structures and a prevalence of short-range connections (Bassett et al., 2010; Kaiser and Hilgetag, 2006; Felleman and Van Essen, 1991). The physical distance between neurons is a critical factor since it determines the material wiring cost, and it increases signal propagation delay, which can impact the timing of neural computations (Buzsáki, 2006). This suggests a fundamental trade-off between minimizing wiring cost and delay (favoring local connections) and global integration and complex computation (favoring long-range connections) (Bassett and Sporns, 2017). Ignoring such spatial constraints will likely result in in computational models with unrealistic connectivity patterns.

Spatial Constraints in Neural Models

Recognizing the importance of physical constraints, various methods have incorporated spatial aspects into neural network models, such as developmental models where network structure emerges through growth rules and spatial constraints (Kaiser et al., 2010). Others use neurons in geometric space and introduce distance-dependent connectivity probabilities or penalties during network construction or learning (Rubinov and Sporns, 2011). Regularization techniques offer a flexible framework for incorporating such constraints during the training of neural networks. Regularization methods like L1 (Lasso) and L2 (Ridge) penalize large weight magnitudes, promoting sparsity or smaller weights, to prevent overfitting (Tibshirani, 1996; LeCun et al., 2015). Others employ spatial-aware regularization (Achterberg et al., 2023) that penalizes recurrent connection weights based on the Euclidean distance between the pre- and post-synaptic neurons embedded in 3D space. Also, this regularizer modulates the wiring cost distance penalty (Estrada and Hatano, 2008), to reflect the functional and communication roles of connections in network dynamics.

Network Size, Scaling, and ALife Principles

Neural network size, the number of neurons (N), is a critical factor influencing computational capacity and learning. Larger networks can represent more complex functions and store more information, but also have higher metabolic and volume costs (in biological systems) and more parameters and training time (in computational systems) (Hinton, 1986). How network size interacts with constraints, such as wiring cost, is an important but little-studied question. For example, efficient processing concomitant with increased network size might necessitate increased modularity as observed in biological brains (Grindrod and Higham, 2018). From an ALife perspective, studying systems under constraints is fundamental (Gershenson, 2023). Investigating how adaptive behavior (task performance) emerges in RNNs under the pressures of task requirements and resource limitations (implicit in size N and explicit via spatial cost λ) aligns with the goal of understanding complex adaptive systems. The trade-offs explored in this study, between network capacity (related to N), connectivity cost (related to λ), and task performance, reflect broader principles of efficiency and adaptation relevant to biological and artificial systems.

Methods

Computational Task: Seasonal Foraging

We developed a *seasonal foraging* task (Nolfi and Parisi, 1996) designed to evaluate an agent's capability to learn and apply context-dependent rules for consuming food items.

Task Structure: Each trial comprises a season cue, a delay and a food cue phase, where an agent (network) trial duration is $T = T_{\text{season}} + T_{\text{delay}} + T_{\text{food}} = 50$ steps.

1. *Season Cue Phase* ($T_{\text{season}} = 20$ steps): The current season is presented to the agent (neural network).
2. *Delay Phase* ($T_{\text{delay}} = 10$ steps): No external cues, and the agent must represent season information internally.
3. *Food Cue Phase* ($T_{\text{food}} = 20$ steps): A specific food item is presented to the agent.

Input Representation: The agent receives sensory input vectors $\boldsymbol{x}_t \in \mathbb{R}^{D_{\text{in}}}$ at each time step t. The input dimension $D_{\text{in}} = D_{\text{season}} + D_{\text{food}}$, where $D_{\text{season}} = 4$ is the number of seasons and $D_{\text{food}} = 6$ is the number of food types (colors), where to enhance task-accomplishing robustness, Gaussian noise ($\mathcal{N}(0, 0.05^2)$) is added to input vectors per time step.

During the season phase ($1 \leq t \leq T_{\text{season}}$), the sensory input is $\boldsymbol{x}_t = [\boldsymbol{s}; \mathbf{0}_{\text{food}}]$, where $\boldsymbol{s}$ is a one-hot vector representing the current season (for example, $[1, 0, 0, 0]^\top$ for Autumn) and $\mathbf{0}_{\text{food}}$ is a zero vector of size D_{food}.

In the delay phase ($T_{\text{season}} < t \leq T_{\text{season}} + T_{\text{delay}}$), the sensory input is a zero vector: $\boldsymbol{x}_t = \mathbf{0}_{\text{in}}$, and in the food phase ($T_{\text{season}} + T_{\text{delay}} < t \leq T$), the agent's input is $\boldsymbol{x}_t = [\mathbf{0}_{\text{season}}; \boldsymbol{f}]$, where $\boldsymbol{f}$ is a one-hot vector representing the food color (for example, $[0, 1, 0, 0, 0, 0]^\top$ for Red) and $\mathbf{0}_{\text{season}}$ is a zero vector of size D_{season}.

Output Representation: At the end of a trial, the agent produces a decision $\boldsymbol{y} \in \mathbb{R}^{D_{\text{out}}}$, where $D_{\text{out}} = 2$. The agent's output represents the probability distribution over two actions: *Eat* ($[1, 0]^\top$) and *Don't Eat* ($[0, 1]^\top$).

Task Rules and Cognitive Demands: Correct decision making depends on a combination of the season and food to the agent presented given predefined rules (Table 1), meaning seasonal foraging task accomplishment necessitates:

1. *Long-Term Memory (LTM):* Encodes associations between seasons and between safe and unsafe foods.
2. *Working Memory (WM):* Keeps the current season's identity across delay periods when no external cue is available.
3. *Perceptual Processing:* Differentiating the input patterns corresponding to distinct seasons and food colors.

Season	Eat	Don't Eat
Autumn	Yellow, Red	Blue, Green, Purple, Black
Winter	Yellow, Blue	Red, Green, Purple, Black
Spring	Yellow, Green	Blue, Red, Purple, Black
Summer	Yellow, Purple	Blue, Green, Red, Black

Table 1: Seasonal foraging task: Which food colors are designated as *Eat* (safe) versus *Don't Eat* (unsafe), per season.

4. *Action Selection:* Integrating the retrieved LTM rules with the current WM state (season) and the perceived food cue to make the appropriate binary decision.
5. *Learning:* Acquiring the LTM rules through experience.

Justification: The seasonal foraging task provides an environment with sufficient dynamics (sensory variation) necessary to evaluate how network structure, given size and spatial constraints, enables task accomplishing behaviors that emerge in response to environmental changes over time. The task serves as a proxy for adaptive behaviors where context (season) modulates responses to stimuli (food) over time, requiring memory and flexible decision-making.

Recurrent Neural Network Model

We use a SimpleRNN architecture implemented in TensorFlow(Keras)[1], where the network architecture consists of:

1. Input layer: time series data: $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$.
2. A Gaussian noise layer.
3. A single recurrent hidden layer comprising N SimpleRNN units, where N is a key experimental parameter varied across $\{50, 100, 200\}$.
4. Output layer: D_{out}=2 units, softmax activation function.

The RNN includes a hidden state $\boldsymbol{h}_t \in \mathbb{R}^N$, where the RNN layer is updated per time step t using equation 1.

$$\boldsymbol{h}_t = \phi(\boldsymbol{W}_{\text{in}}\boldsymbol{x}_t + \boldsymbol{W}_{\text{rec}}\boldsymbol{h}_{t-1} + \boldsymbol{b}_{\text{rec}}) \tag{1}$$

Where, $\boldsymbol{W}_{\text{in}} \in \mathbb{R}^{N \times D_{\text{in}}}$ are the input weights, $\boldsymbol{W}_{\text{rec}} \in \mathbb{R}^{N \times N}$ are the recurrent weights, $\boldsymbol{b}_{\text{rec}} \in \mathbb{R}^N$ is the bias vector, and ϕ is the activation function. We use the Rectified Linear Unit (ReLU) activation function: $\phi(z) = \max(0, z)$. Recurrent weights $\boldsymbol{W}_{\text{rec}}$ are initialized using an orthogonal matrix scheme. For the final classification decision, only the last hidden state $\boldsymbol{h}_T$ is passed to the output layer. In the output layer the final output $\boldsymbol{y}$ is computed using equation 2.

$$\boldsymbol{y} = \text{softmax}(\boldsymbol{W}_{\text{out}}\boldsymbol{h}_T + \boldsymbol{b}_{\text{out}}) \tag{2}$$

Where, $\boldsymbol{W}_{\text{out}} \in \mathbb{R}^{D_{\text{out}} \times N}$ are the output weights and $\boldsymbol{b}_{\text{out}} \in \mathbb{R}^{D_{\text{out}}}$ is the output bias.

[1] https://www.tensorflow.org/guide/keras

Spatial Embedding, Distance-based Regularization

A key component of our approach are spatial constraints on the network's connectivity, implemented via a custom regularization term applied to the recurrent weight matrix $\boldsymbol{W}_{\text{rec}}$.

Neuron Coordinates: We assume the N neurons of the RNN layer are physically embedded in a 3D Euclidean space. For each network size N, we generate a set of coordinates $\{\boldsymbol{c}_i \in \mathbb{R}^3\}_{i=1}^{N}$. These coordinates are derived by defining a minimal cubic grid structure large enough to contain N points, populating the grid, and selecting the first N coordinates. This procedure ensures a consistent spatial layout across network sizes. The Euclidean distance between neuron i and j is denoted by $d_{ij} = ||\boldsymbol{c}_i - \boldsymbol{c}_j||_2$, where distances are stored in a distance matrix $\boldsymbol{D} = [d_{ij}]$.

Communicability-Weighted Spatial Regularization: We introduce a regularization penalty $R(\boldsymbol{W}_{\text{rec}})$ added to the network's loss function during training. This encourages sparsity and locality by penalizing recurrent connection magnitude $|\boldsymbol{W}_{\text{rec},ij}|$ based on the physical distance d_{ij} between connected neurons, modulated by a connection's role in network-wide communication, as outlined in following.

1. *Base Spatial Cost:* The fundamental cost of a connection is proportional to its absolute weight and the distance.
2. *Communicability Modulation:* To account for the functional importance of connections within the network dynamics, we compute a communicability matrix $\boldsymbol{G}$. First, the absolute weight matrix $|\boldsymbol{W}_{\text{rec}}|$ is normalized to obtain $\boldsymbol{W}_{\text{norm}} = (\text{diag}(\boldsymbol{s}))^{-1/2}|\boldsymbol{W}_{\text{rec}}|(\text{diag}(\boldsymbol{s}))^{-1/2}$, where $\boldsymbol{s}$ is the vector of node strengths ($s_i = \sum_j |\boldsymbol{W}_{\text{rec},ij}|$) and $\text{diag}(\boldsymbol{s})$ is the diagonal matrix of strengths (with a small epsilon added to s_i for numerical stability (Rajapandian et al., 2020)). The communicability matrix is then calculated as $\boldsymbol{G} = \text{expm}(\boldsymbol{W}_{\text{norm}})$, where expm denotes the matrix exponential. $\boldsymbol{G}_{ij}$ quantifies the ease of communication between nodes i and j through all possible paths within the network, thus accounting for a connection's role in facilitating network-wide information flow beyond its direct weight or distance (Estrada and Hatano, 2008).
3. *Penalty:* Equation 3 defines the final regularization term.

$$R(\boldsymbol{W}_{\text{rec}}) = \lambda \sum_{i=1}^{N}\sum_{j=1}^{N} |\boldsymbol{W}_{\text{rec},ij}| \cdot d_{ij} \cdot \boldsymbol{G}_{ij} \tag{3}$$

Where, λ is the regularization strength parameter. This penalty discourages connections that are simultaneously long, strong, and part of highly communicable pathways.

Training and Statistical Analysis

Learning Paradigm: Per trial, the network is input with the sequence: $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$ and generates an output $\boldsymbol{y}$. The objective is to minimize the difference between output and the target decision $\boldsymbol{y}_{\text{target}}$ (one-hot encoded correct action).

Loss Function: Equation 4 defines the standard categorical cross-entropy loss function:

$$L = -\sum_{k=1}^{D_{\text{out}}} y_{\text{target},k} \log(y_k) \tag{4}$$

Optimization: The total loss minimized during training is $L_{\text{total}} = L + R(\boldsymbol{W}_{\text{rec}})$, performed using the Adam optimizer (Diederik, 2014) with default parameters.

Data Generation and Training Regime: Per training run, $N_{\text{train}} = 10240$ trials are generated. Training proceeds for 100 epochs, with inputs shuffled and presented in batches of size 128. A separate validation set ($N_{\text{val}} = 5120$ trials) is used to generalize task performance during training.

Non-parametric tests were used to assess statistical significance differences in task performance and structural metrics between network sizes (N) per regularization strength (λ). Per metric and each fixed λ value, we used the Kruskal-Wallis H test (Kruskal and Wallis, 1952) to determine if there was a significant difference among the median values of the three network size groups based on the N_{runs}=20 runs per treatment. If the Kruskal-Wallis test yielded a significant result ($p < 0.05$), we performed Dunn's post-hoc test with Bonferroni correction (Dunn, 1964) for multiple comparisons to identify which pairs of network sizes differed significantly. Test statistics and p-values are online[2].

Experimental Design and Data

Experiments systematically varied two key parameters: the network size $N \in [50, 100, 200]$ and the regularization strength $\lambda \in [0.01, ..., 0.25]$. For each combination of (N, λ), we performed $N_{\text{runs}} = 20$ independent training runs, each starting with different random weight initializations and training data order, to ensure statistical robustness of the results. This enabled us to analyze how the interaction between intrinsic network scale (N) and extrinsic spatial constraints λ shapes the structure, dynamics, and task performance of the learned recurrent networks.

Experimental Setup

Experiments evaluated the impact of network size (N) and the strength of spatial regularization (λ) on task performance and emergent structure of RNNs trained on the seasonal foraging task. Experiments varied the number of recurrent units (N) and the regularization hyperparameter (λ). All other architectural, task, and training parameters

[2]Source code and results: `https://anonymous.4open.science/r/Spacially-aware-RNN`

Table 2: Key Experimental Parameters

Parameter	Value(s)
Independent Variables	
Network Size (Recurrent Units), N	$\{50, 100, 200\}$
Regularization Strength, λ	$\{0.01, ..., 0.25\}$
Task Parameters	
Number of Seasons, D_{season}	4
Number of Food Types, D_{food}	6
Season Cue Duration, T_{season}	20 steps
Delay Duration, T_{delay}	10 steps
Food Cue Duration, T_{food}	20 steps
Input Noise, σ	0.05
Training Parameters	
Epochs	100
Batch Size	128
Training Trials per Run, N_{train}	10240
Validation/Test Trials, $N_{\text{val/test}}$	5120
Optimizer	Adam
RNN Activation Function, ϕ	ReLU
Recurrent Initializer	Orthogonal
Regularization Parameters	
Distance Metric	Euclidean
Experimental Procedure	
Independent Runs per Condition, N_{runs}	20

(Table 2) were held constant across experiments in order to isolate the impact of these two key variables.

Network Size (N): The range $\{50, 100, 200\}$ was chosen to span between small networks, where resource limitations might be significant, to larger networks capable of more complex solutions. This enables us to observe how scaling impacts task performance given spatial constraints.

Regularization Strength (λ): The range was a spectrum from weak spatial influence (close to an unregularized network) to strong spatial influence (distance penalty significantly impacts weight formation). The non-linear spacing of values aims to capture potentially rapid transitions in network structure or performance as the penalty increases.

Number of Runs (N_{runs}): We performed 20 runs per (N, λ) treatment to ensure statistical robustness, accounting for the effects of random weight initialization and stochasticity in the training process (for example, batch sampling).

Epochs and Trial Counts: 100 epochs were sufficient for networks to converge to stable task performance. The number of training trials (N_{train}=10240), validation and test set size (N_{test}=5120) enabled reliable task evaluation.

Task Duration: Durations for season, delay, and food phases defined the periods requiring perceptual processing ($T_{\text{season}}, T_{\text{food}}$) and internal maintenance (T_{delay}), ensuring the task is suitable to test our research questions.

Results and Discussion

Experiment results report the impact of network size (N) and communicability-weighted spatial regularization strength (λ) on the agent's RNN emergent structural properties as well as seasonal foraging task capability.

Network Size Modulates Tolerance to Spatial Costs

We first examined how task performance (final accuracy on the test set) varied with network size N and regularization strength λ. Accuracy represents the fraction of correctly classified trials, ranging from 0.0 (low performance) to 1.0 (perfect performance), with higher values indicating better task execution (Figure 1). All network sizes were capable of achieving high average accuracy (≥ 0.90) on the seasonal foraging task, demonstrating the general general efficacy of our selected neural architecture (Figure 1).

Performance sensitivity to the regularization strength λ differed visually between network sizes (Figure 1). While all networks achieved perfect or near-perfect peak accuracy at low-to-moderate penalties, their degradation patterns varied. The smallest network ($N = 50$, peak at $\lambda = 0.03$) exhibited relatively graceful degradation, maintaining performance above 0.900 throughout. Similarly, the $N = 100$ network (peak at $\lambda = 0.05$) also degraded gradually. In contrast, the largest network ($N = 200$, peak at $\lambda = 0.01$) was more sensitive, with a sharper drop reaching the lowest minimum average accuracy (0.883 at $\lambda = 0.16$) despite recovering to the highest final accuracy at $\lambda = 0.25$ (0.913). However, these observed differences were not statistically significant at any single λ level (Kruskal-Wallis, $p > 0.05$ for all λ). The lack of statistical significance, despite differing degradation profiles, is posited to result from the task allowing for multiple solutions, leading to a ceiling effect for this task performance metric across networks with adequate capacity (Barak and Tsodyks, 2014).

This task performance variation suggests larger networks, despite greater capacity, might adopt architectures more reliant on specific long-range connections, rendering them vulnerable when such connections are heavily penalized. This resonates with observations in mammalian cortical scaling, where increases in cortical area necessitate a disproportionate increase in the volume of costly, long-range white matter tracts (Chklovskii et al., 2002; Bullmore and Sporns, 2012). Our results illustrate this capacity-cost tradeoff, where connections required for attaining some minimal task performance are adapted for given spatial constraints.

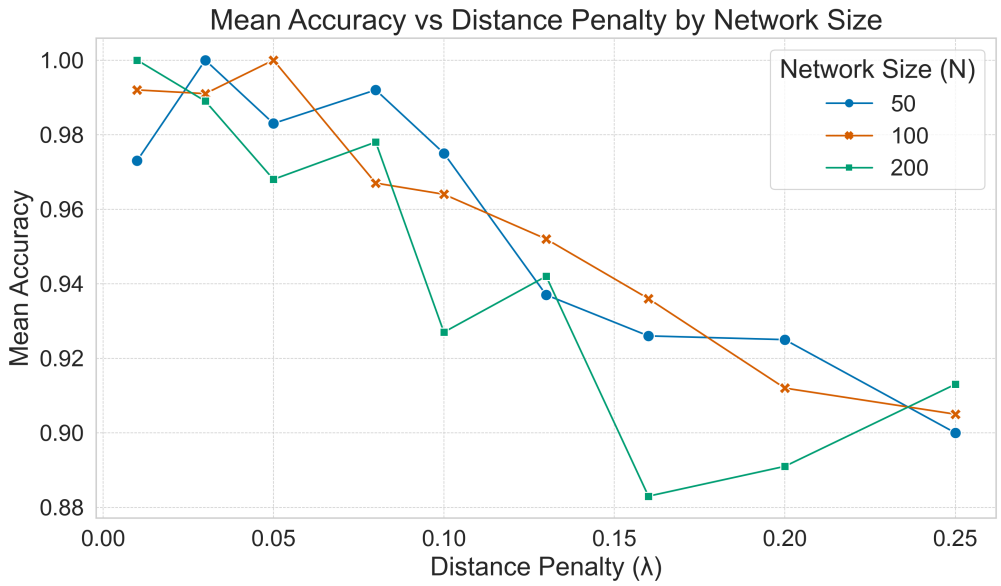

Figure 1: Task Performance (Accuracy) versus regularization strength (λ) across network sizes (N). Mean accuracy across N_{runs}=20 runs is shown, illustrating high peak performance is achievable across network sizes.

Emergence of Brain-Like Network Topologies

Next we analyzed the structural properties of the learned recurrent weight matrices ($\boldsymbol{W}_{rec}$) to understand how spatial regularization shapes the agent's neural network topology.

Connection Length and Sparsity: Observing network graphs[2] revealed a clear trend: increasing λ favored shorter connections, yet even under strong penalties (λ=0.25), a few long-range connections were typically preserved. Quantitatively, median connection length decreased monotonically with λ across all network sizes. Retained long-range connections are theorized to represent functionally critical pathways analogous to sparse but vital long-distance *fasciculi* in the primate brain (Bullmore and Sporns, 2012). Concurrently, the total recurrent weight magnitude ($\sum_{i,j} |W_{rec,ij}|$), indicating overall connectivity cost and density, decreased steeply with increasing λ before plateauing or slightly increasing at extreme penalties (Figure 2), reflecting compensatory strengthening of essential local connections. Due to scale, total weight differed significantly between network sizes at all λ values (Kruskal-Wallis, p < 0.0001), with all pairwise comparisons being significant (Dunn's test, p < 0.05), confirming that larger networks maintain a significantly greater magnitude of connections.

Modularity: We quantified network modularity (Q) using the Louvain algorithm (Blondel et al., 2008) (Figure 3), where higher positive values indicate stronger segregation into communities, a hallmark of organization in many complex biological brain networks (Bassett et al., 2010; Sporns et al., 2004). Modularity increased significantly with λ for larger networks ($N = 100, 200$), indicating penalty-driven formation of functional clusters (Q $\approx$ 0.15-0.18). Interestingly, the small $N = 50$ network exhibited significantly higher baseline modularity (Q $\approx$ 0.26) compared to larger networks at low penalties ($\lambda \leq 0.05$; Kruskal-Wallis, p < 0.01; Dunn's test, p < 0.01 for $N = 50$ versus $N = 100, N = 200$) and followed a non-monotonic, U-shaped trend. Significant differences between sizes persisted at specific higher penalties (λ=0.10, 0.16, 0.25; Kruskal-Wallis, p < 0.05), often with $N = 50$ remaining distinct (Dunn's test, p < 0.05). This suggests size-dependent emergent structure, indicating ceiling effects or topological phases in smaller versus larger systems given spatial constraints (Fornito et al., 2010).

Small-Worldness: We computed the small-world coefficient σ (Watts and Strogatz, 1998; Humphries and Gurney, 2008), where $\sigma > 1$ is an efficient topology balancing local clustering and global integration, characteristic of biological neural systems (Bassett et al., 2010; Sporns et al., 2004). All networks displayed small-world properties ($\sigma > 1$. Figure 4), with σ generally increasing with λ, especially for $N = 50$. Significant differences in σ between network sizes emerged primarily at intermediate penalties (λ=0.05, 0.10, 0.13; Kruskal-Wallis, p < 0.05). For instance, $N = 50$ was significantly higher σ than $N = 200$ at λ=0.05, while $N = 100$ differed significantly from $N = 200$ at λ=0.10 and λ=0.13 (Dunn's test, p < 0.05). Significant differences in this intermediate regime indicate a critical point where

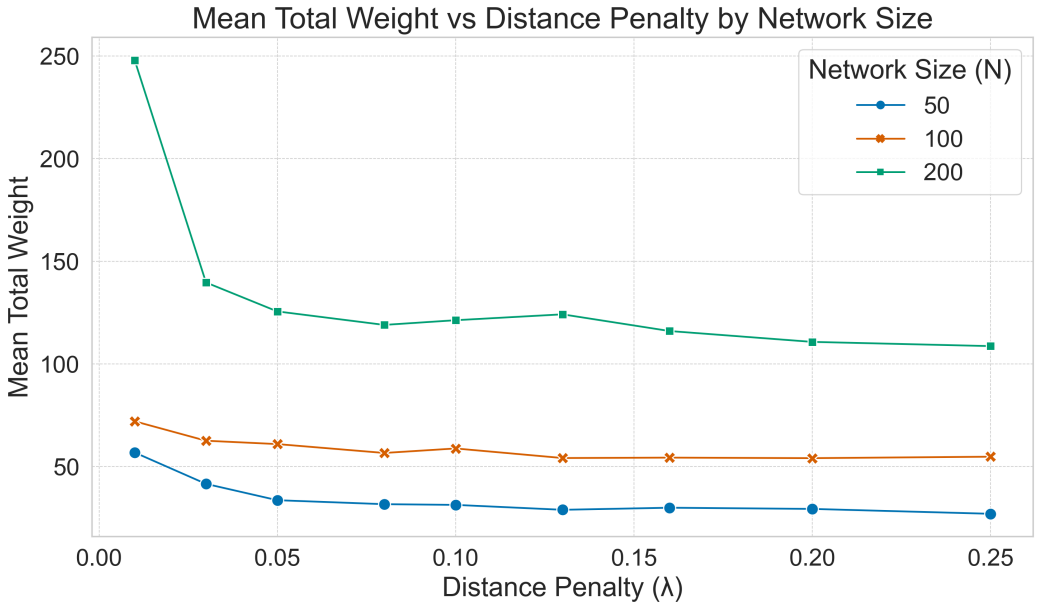

Figure 2: Mean total recurrent weight ($\sum |\boldsymbol{W}_{rec,ij}|$) versus regularization strength (λ) across network sizes (N). Decreasing weight indicates sparsification due to λ. The scale differs significantly between sizes ($p<0.0001$), given capacity differences.

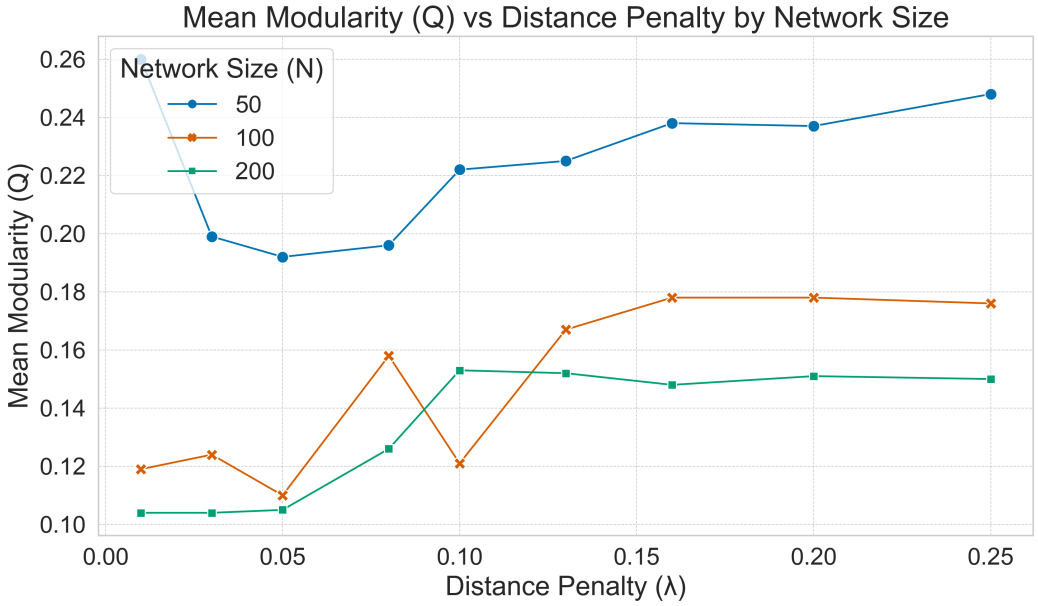

Figure 3: Mean network modularity (Q) versus regularization strength (λ) across network sizes (N). Higher Q indicates stronger community structure.

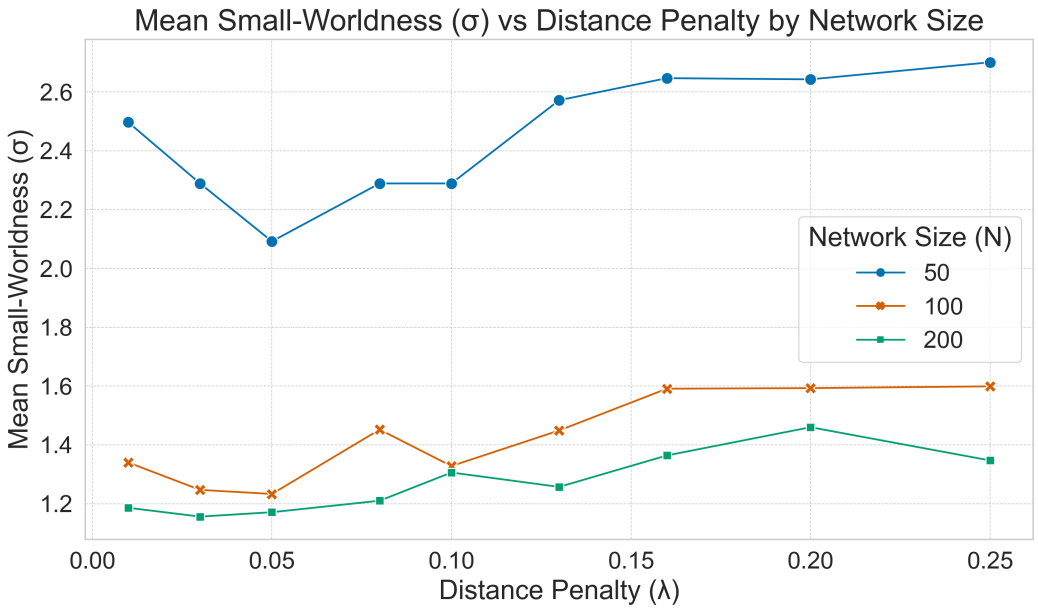

Figure 4: Mean *small-worldness* coefficient (σ) versus regularization strength (λ) across network sizes (N): $\sigma > 1$ indicates small-world characteristics.

spatial cost actively forces localization, making the resulting balance (and thus σ) most sensitive to network size (N) before high penalties dominate (Watts and Strogatz, 1998).

Network Size-Cost and Learning Dynamics

The regularization strength λ influenced network structure and learning dynamics. Analysis over epochs[2] indicated a two-phase dynamic for strong penalties ($\lambda \geq 0.10$): rapid initial pruning and gradual stabilization. Low-λ networks learned faster than high-λ networks, showing a trade-off between convergence speed and finding cost-efficient solutions. This dynamic is similar to developmental processes in the mammalian cortex (Huttenlocher, 1990).

These differing dynamics result in the observed interaction between network capacity (N) and wiring cost penalty (λ). The results highlight a crucial trade-off: while final accuracy plateaus across sizes (Figure 1), the underlying network cost (total weight, Figure 2) remains significantly different, and the emergence of efficient structural properties (modularity, small-worldness, Figures 3 and 4) is most pronounced at intermediate λ values that still permit high task accuracy (≥ 0.95). This suggests an optimal regime (*sweet spot*) where biologically plausible structural organization can be achieved without significant functional cost, provided the spatial pressure (λ) is appropriately moderated relative to the network's scale (N). These findings directly address our research objectives by quantifying the size-performance interaction (Objective 1), demonstrating the influence of λ on emergent structure (Objective 2), and revealing the size-dependent scaling of these effects (Objective 3). This study's primary contribution is demonstrating that biologically plausible wiring costs can drive self-organization towards efficient, brain-like network topologies in functional RNNs, with the optimal balance between cost and task performance being contingent on network scale.

Conclusion

This study demonstrated that a connection wiring cost coupled with communicability-weighted spatial regularization, guides an RNN towards sparse, modular, and small-world topologies reminiscent of biological brains. Our main finding is a critical interaction between network scale and spatial constraints. Optimal regularization strength (λ) fosters efficient structural properties without sacrificing high task performance. Such optimal regularization scales inversely with network size (N). Thus, when spatial penalties are suitably tuned to network capacity, one can achieve an effective compromise by preserving a subset of long-range connections crucial for global function given predominant short-range connections. These results highlight the fundamental role of physical constraints in shaping the structure and complexity of neural systems (Chklovskii et al., 2002; Chechik et al., 1998), while providing future work guidelines for designing neural systems bound by changing resource (energy) (Nagar et al., 2019) and morphological (embodied) (Watson and Nitschke, 2015; Mailer et al., 2021) constraints.

References

Achterberg, J. et al. (2023). Spatially Embedded Recurrent Neural Networks Reveal Widespread Links Between Structural and Functional Neuroscience Findings. *Nature Machine Intelligence*, 5(12):1369–1381.

Azevedo, F. et al. (2009). Equal Numbers of Neuronal and Nonneuronal Cells make the Human Brain an Isometrically Scaled-up Primate Brain. *Journal of Comparative Neurology*, 513(5):532–541.

Barak, O. (2017). Recurrent Neural Networks as Versatile Tools of Neuroscience Research. *Current opinion in neurobiology*, 46:1–6.

Barak, O. and Tsodyks, M. (2014). Working Models of Working Memory. *Current Opinion in Neurobiology*, 25:20–24.

Bassett, D. et al. (2010). Efficient Physical Embedding of Topologically Complex Information Processing Networks in Brains and Computer Circuits. *PLoS Computational Biology*, 6(4):e1000748.

Bassett, D. and Sporns, O. (2017). Network Neuroscience. *Nature Neuroscience*, 20(3):353–364.

Blondel, V. et al. (2008). Fast Unfolding of Communities in Large Networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008.

Bullmore, E. and Sporns, O. (2012). The Economy of Brain Network Organization. *Nature Reviews Neuroscience*, 13(5):336–349.

Buzsáki, G. (2006). *Rhythms of the Brain*. Oxford University Press, Oxford, United Kingdom.

Chechik, G., Meilijson, I., and Ruppin, E. (1998). Synaptic Pruning in Development: A Computational Account. *Neural Computation*, 10(7):1759–1777.

Chklovskii, D., Schikorski, T., and Stevens, C. (2002). Wiring Optimization in Cortical Circuits. *Neuron*, 34(3):341–347.

Cho, K. et al. (2014). Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 1724–1734, Doha, Qatar. Association for Computational Linguistics.

Diederik, K. (2014). Adam: A Method for Stochastic Optimization. *arXiv:1412.6980*.

Dunn, O. (1964). Multiple Comparisons using Rank Sums. *Technometrics*, 6(3):241–252.

Elman, J. (1991). Distributed Representations, Simple Recurrent Networks, and Grammatical Structure. *Machine Learning*, 7:195–225.

Estrada, E. and Hatano, N. (2008). Communicability in Complex Networks. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 77(3):036111.

Felleman, D. and Van Essen, D. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1(1):1–47.

Fornito, A., Zalesky, A., and Bullmore, E. (2010). Network Scaling Effects in Graph Analytic Studies of Human Resting-state FMRI Data. *Frontiers in systems neuroscience*, 4:1373.

Gershenson, C. (2023). Emergence in Artificial Life. *Artificial Life*, 29(2):153–167.

Greff, K. et al. (2016). LSTM: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, 28(10):2222–2232.

Grindrod, P. and Higham, D. (2018). High Modularity Creates Scaling Laws. *Scientific Reports*, 8(1):9737.

Hinton, G. (1986). Learning Distributed Representations of Concepts. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 8.

Hopfield, J. (1982). Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558.

Humphries, M. and Gurney, K. (2008). Network 'Small-worldness': A Quantitative Method for Determining Canonical Network Equivalence. *PloS one*, 3(4):e0002051.

Huttenlocher, P. (1990). Morphometric Study of Human Cerebral Cortex Development. *Neuropsychologia*, 28(6):517–527.

Kaiser, M. et al. (2010). Neural Signatures of Autism. *Proceedings of the National Academy of Sciences*, 107(49):21223–21228.

Kaiser, M. and Hilgetag, C. (2006). Nonoptimal Component Placement, but Short Processing Paths, Due to Long-distance Projections in Neural Systems. *PLoS Computational Biology*, 2(7):e95.

Kandel, E. et al. (2000). *Principles of Neural Science*, volume 4. McGraw-hill, New York, USA.

Kruskal, W. and Wallis, W. (1952). Use of Ranks in One-criterion Variance Analysis. *Journal of the American statistical Association*, 47(260):583–621.

Laughlin, S. et al. (1998). The Metabolic Cost of Neural Information. *Nature Neuroscience*, 1(1):36–41.

Laughlin, S. and Sejnowski, T. (2003). Communication in Neuronal Networks. *Science*, 301(5641):1870–1874.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep Learning. *Nature*, 521(7553):436–444.

Mailer, C., Nitschke, G., and Raw, L. (2021). Evolving Gaits for Damage Control in a Hexapod Robot. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 146–153, Lille, France. ACM.

Nagar, D., Furman, A., and G., N. (2019). The Cost of Big Brains in Groups. In *Proceedings of the 2019 Conference on Artificial Life*, pages 404–411, Newcastle, United Kingdom. MIT Press.

Nolfi, S. and Parisi, D. (1996). Learning to Adapt to Changing Environments in Evolving Neural Networks. *Adaptive Behavior*, 5(1):75–98.

Rajapandian, M. et al. (2020). Uncovering Differential Identifiability in Network Properties of Human Brain Functional Connectomes. *Network Neuroscience*, 4(3):698–713.

Rubinov, M. and Sporns, O. (2011). Weight-conserving Characterization of Complex Functional Brain Networks. *Neuroimage*, 56(4):2068–2079.

Sporns, O. et al. (2004). Organization, Development and Function of Complex Brain Networks. *Trends in Cognitive Sciences*, 8(9):418–425.

Striedter, G. (2004). *Principles of Brain Evolution*. Oxford University Press, OXford, United Kingdom.

Sutskever, I., Vinyals, O., and Le, Q. (2014). Sequence to Sequence Learning with Neural Networks. In *Proceedings of the 28th International Conference on Neural Information Processing Systems*, pages 3104–3112, Montreal, Canada. MIT Press.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288.

Wang, X. (2002). Probabilistic Decision Making by Slow Reverberation in Cortical Circuits. *Neuron*, 36(5):955–968.

Watson, J. and Nitschke, G. (2015). Evolving Robust Robot Team Morphologies for Collective Construction. In *Proceedings of the IEEE Symposium Series on Computational Intelligence*, pages 1039–1046, Cape Town, South Africa. IEEE Press.

Watts, D. and Strogatz, S. (1998). Collective Dynamics of 'Small-world' Networks. *Nature*, 393(6684):440–442.

Yang, G. et al. (2019). Task Representations in Neural Networks Trained to Perform many Cognitive Tasks. *Nature Neuroscience*, 22(2):297–306.