

# Artificial Neural Networks Functionalized By Evolution

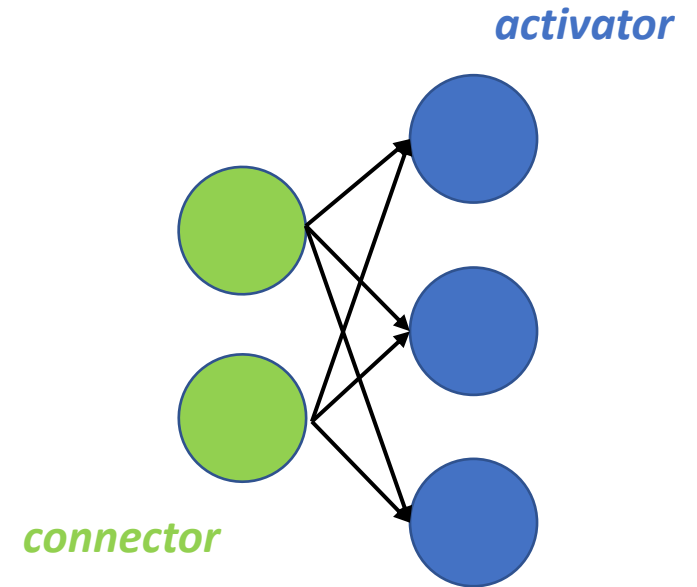
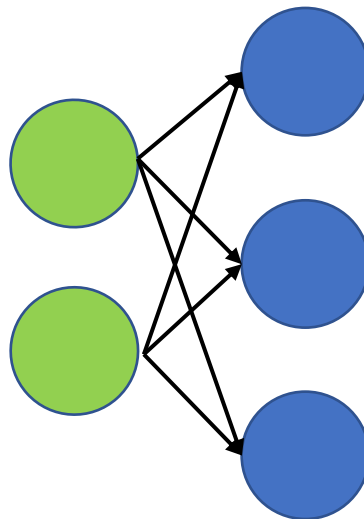
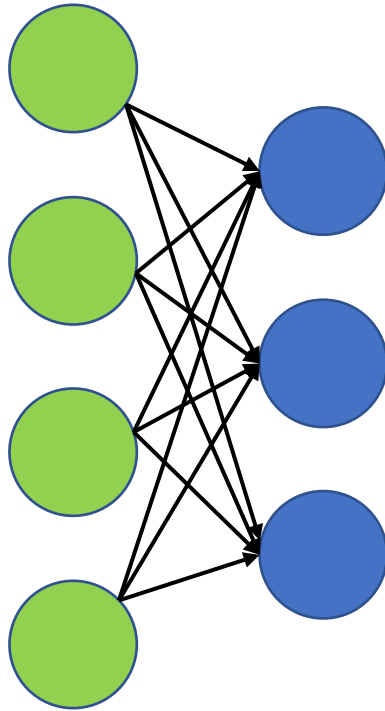
Furfaro, Fabien, Avner Bar-Hen, and Geoffroy Berthelot. "An Artificial Neural Network Functionalized by Evolution." *arXiv preprint arXiv:2205.10118* (2022).

Presented By  
Dennis Manjaly Joshy, ME  
Yi Lab, UCSB

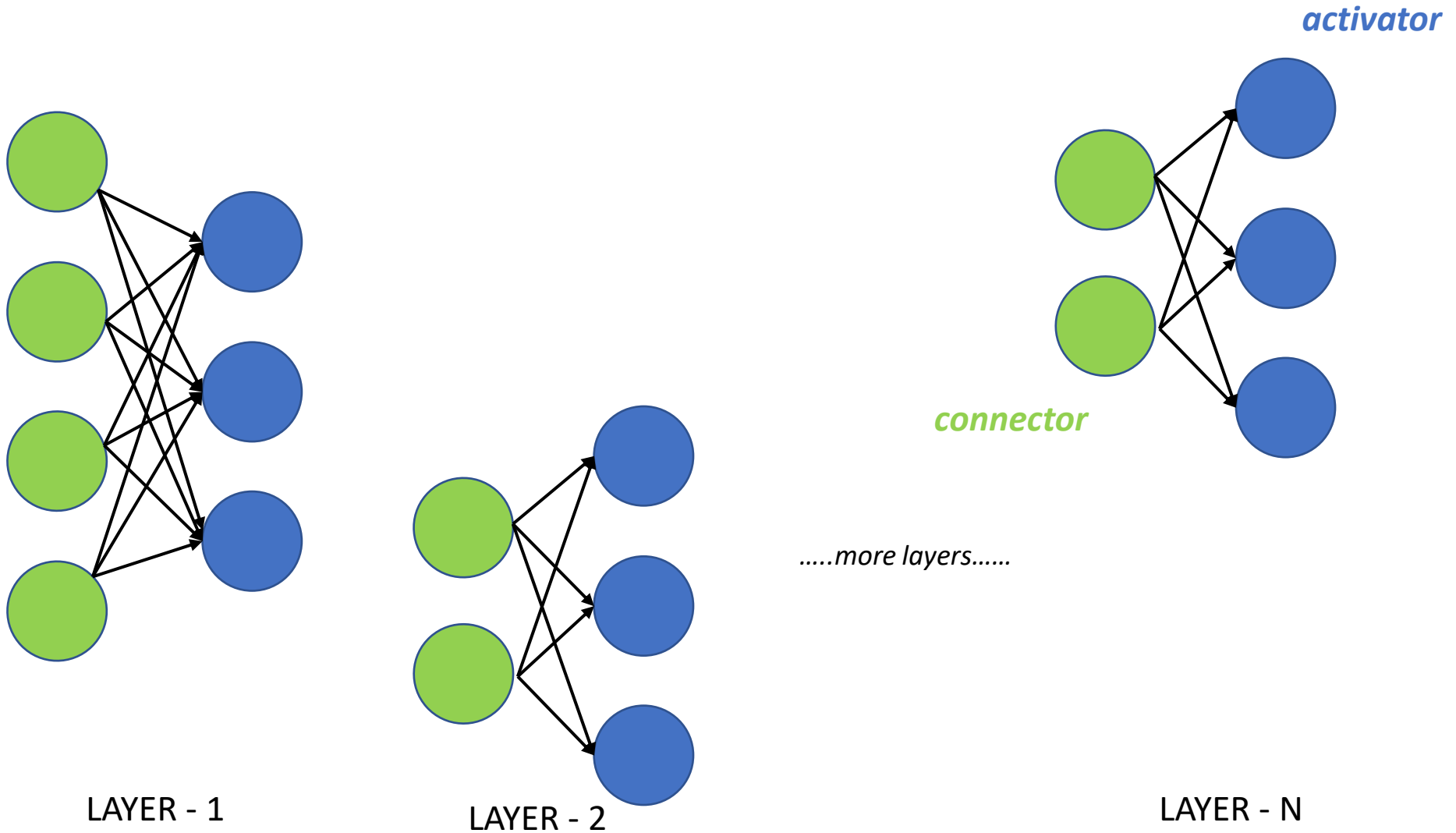
# Outline

- Results and Discussion from Furfaro et. Al (2022)
- Hurwitz Stability Informed Evolutionary Learning

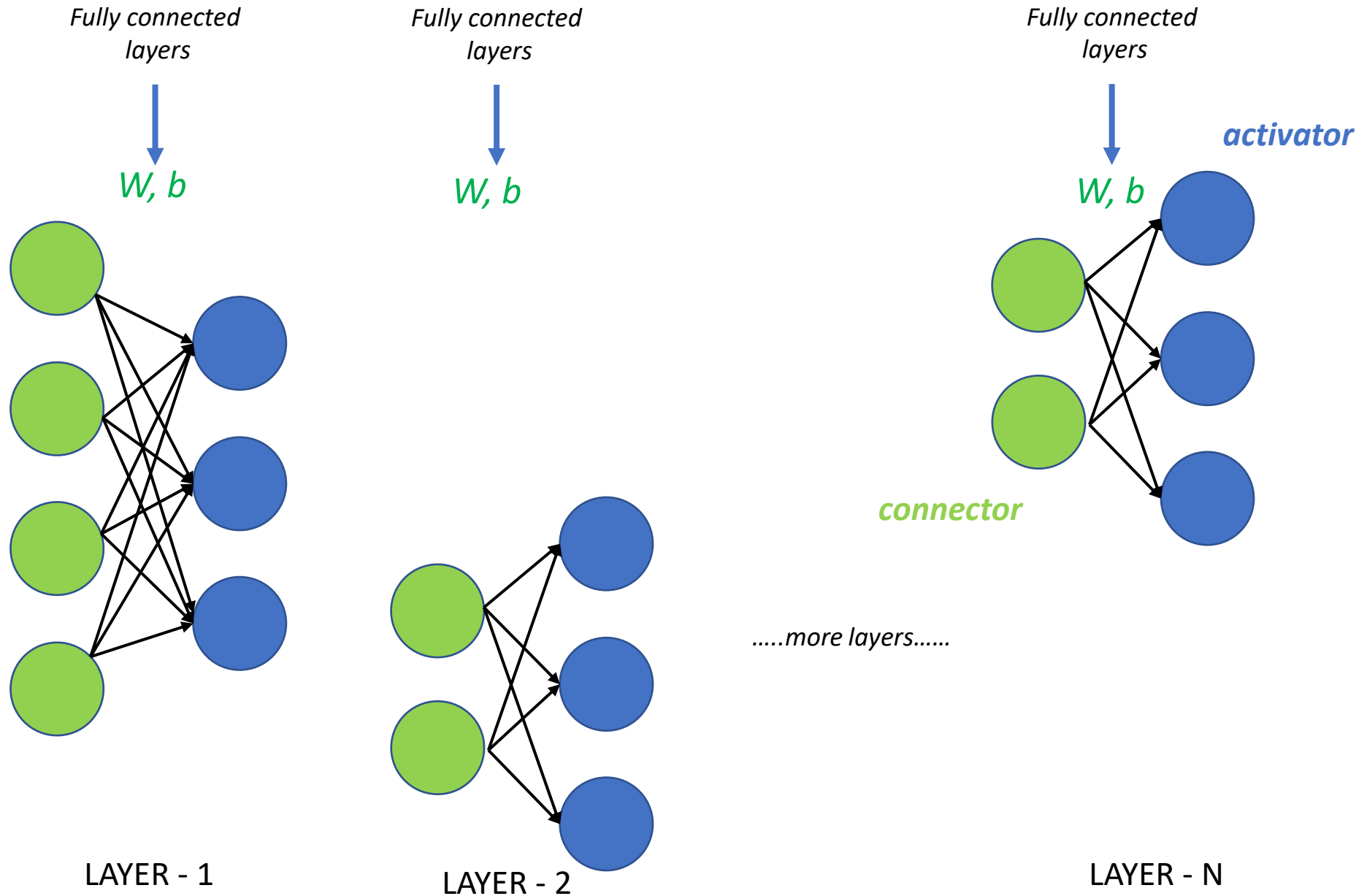
# *A more 'Neuro-Inspired' Topology*



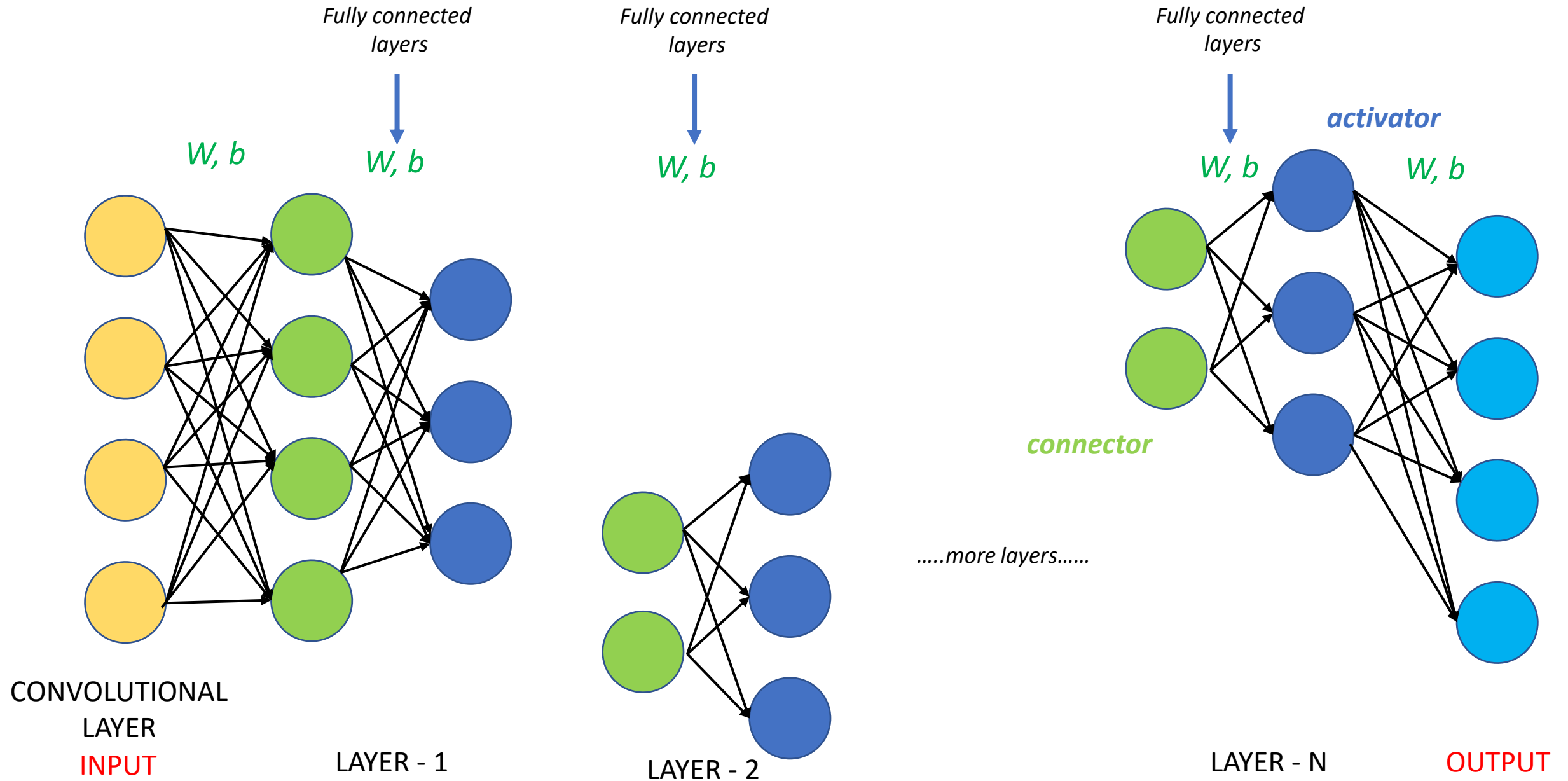
# A more 'Neuro-Inspired' Topology



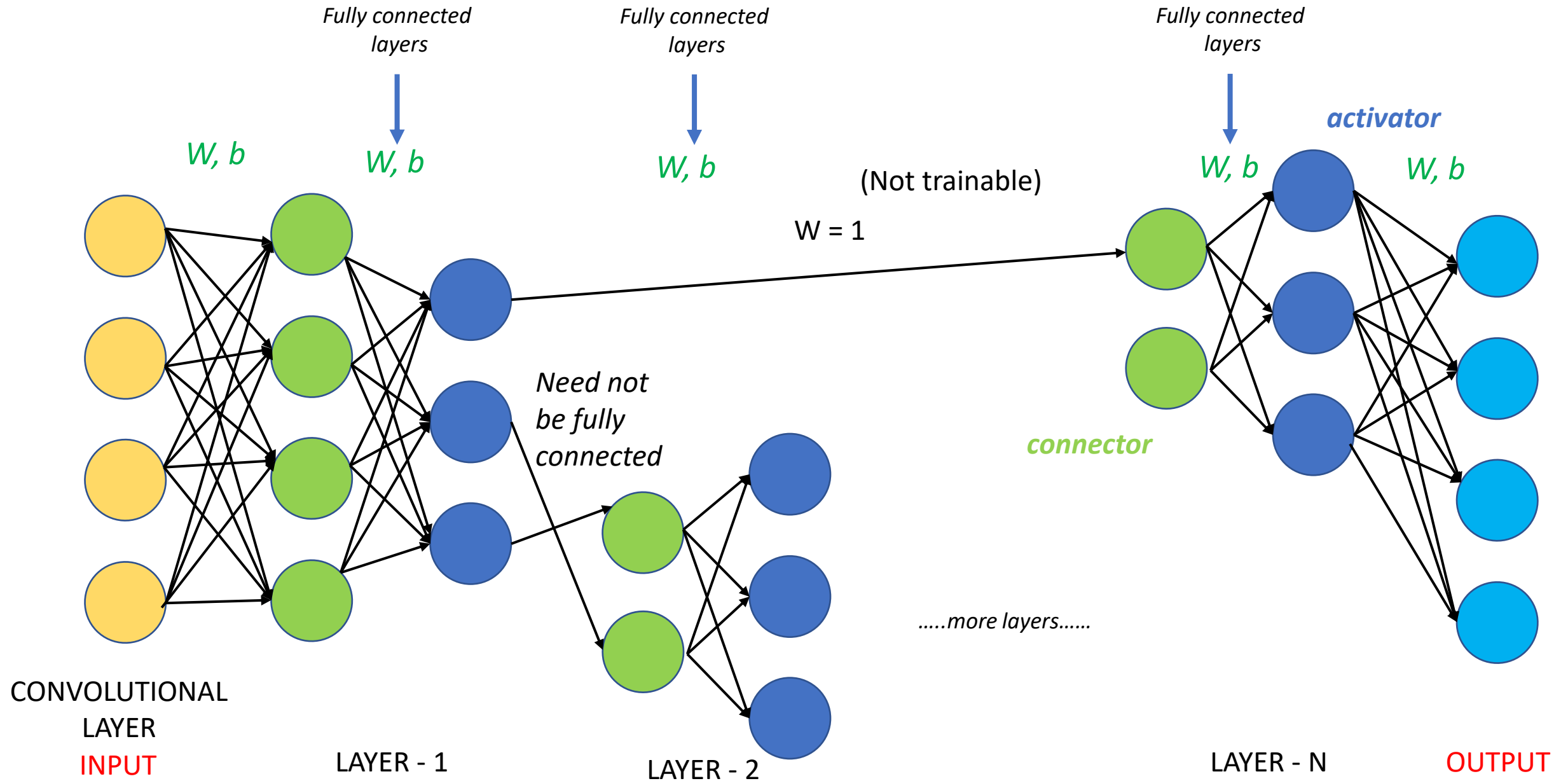
# A more 'Neuro-Inspired' Topology



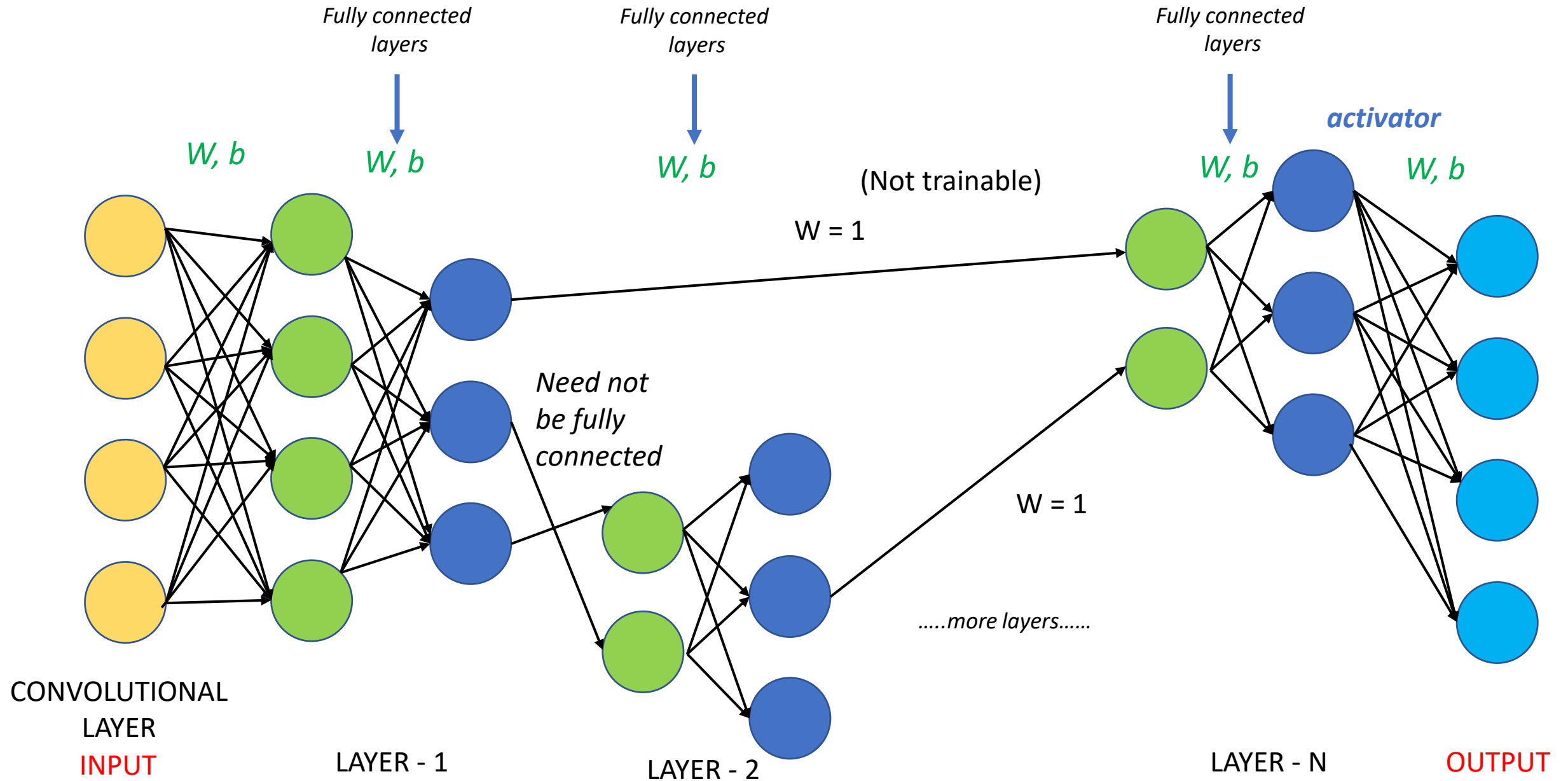
# A more 'Neuro-Inspired' Topology



# A more 'Neuro-Inspired' Topology



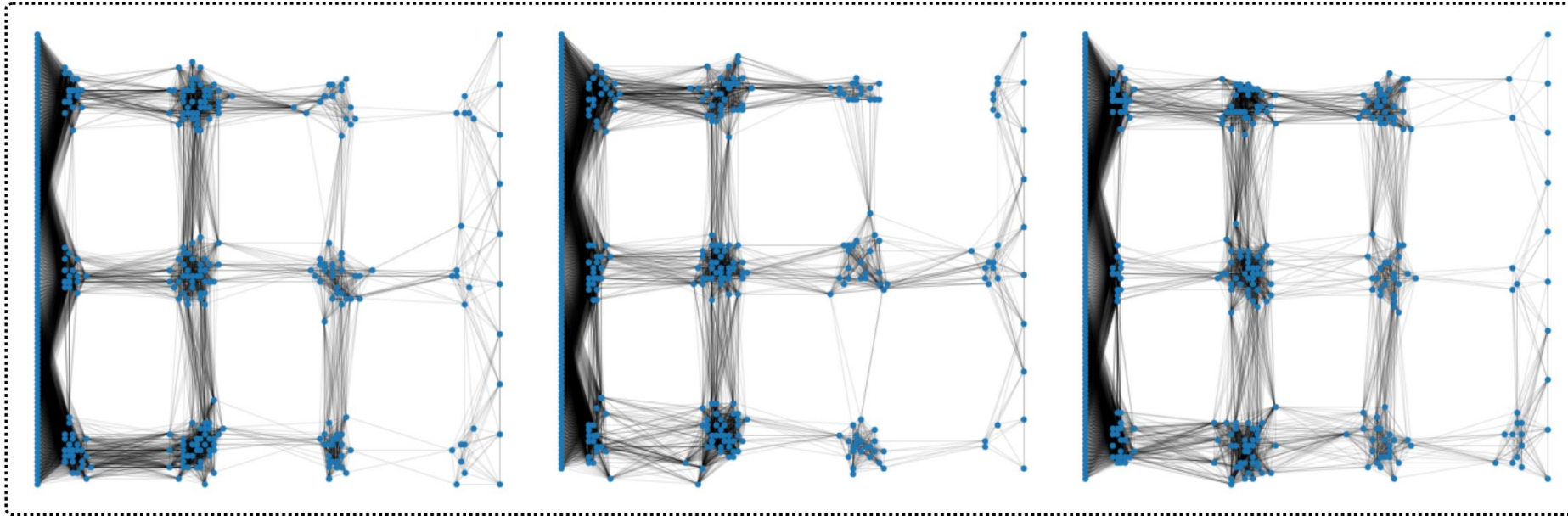
# A more 'Neuro-Inspired' Topology





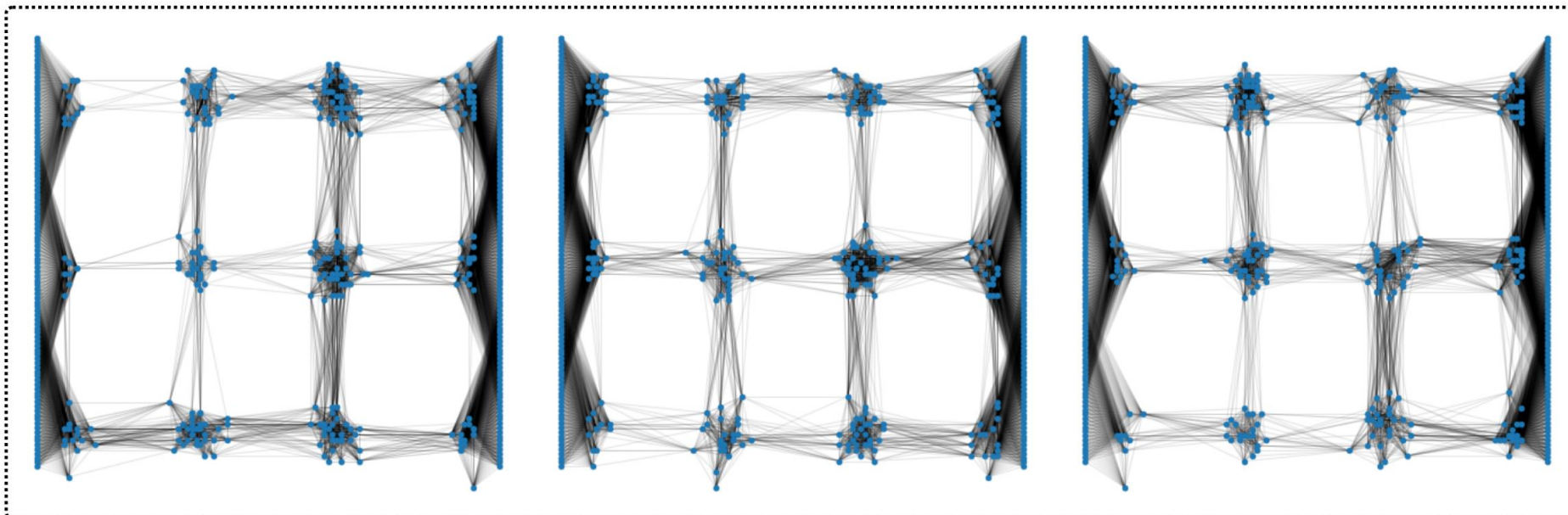
# Visualizing the Networks Made with these rules

Taille In  $\neq$  Out



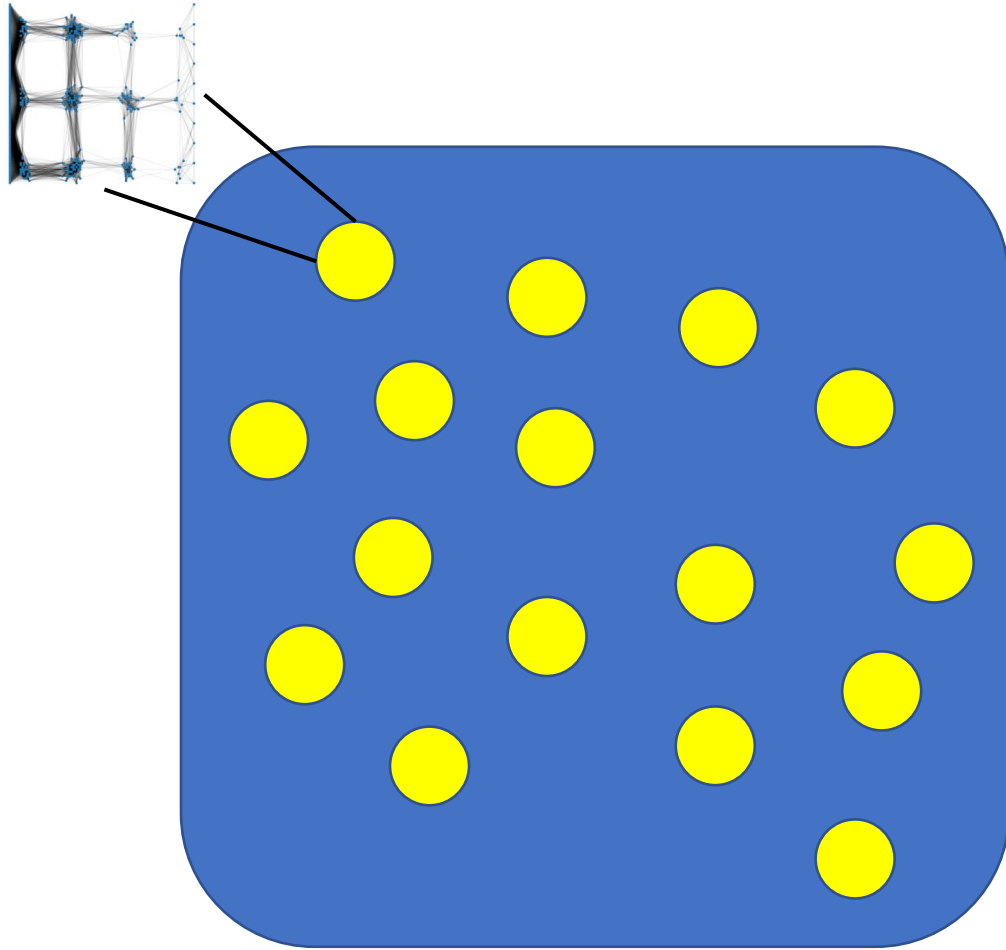
Applications :  
Base MNIST / IMdb  
Prise de décision

Taille In = Out



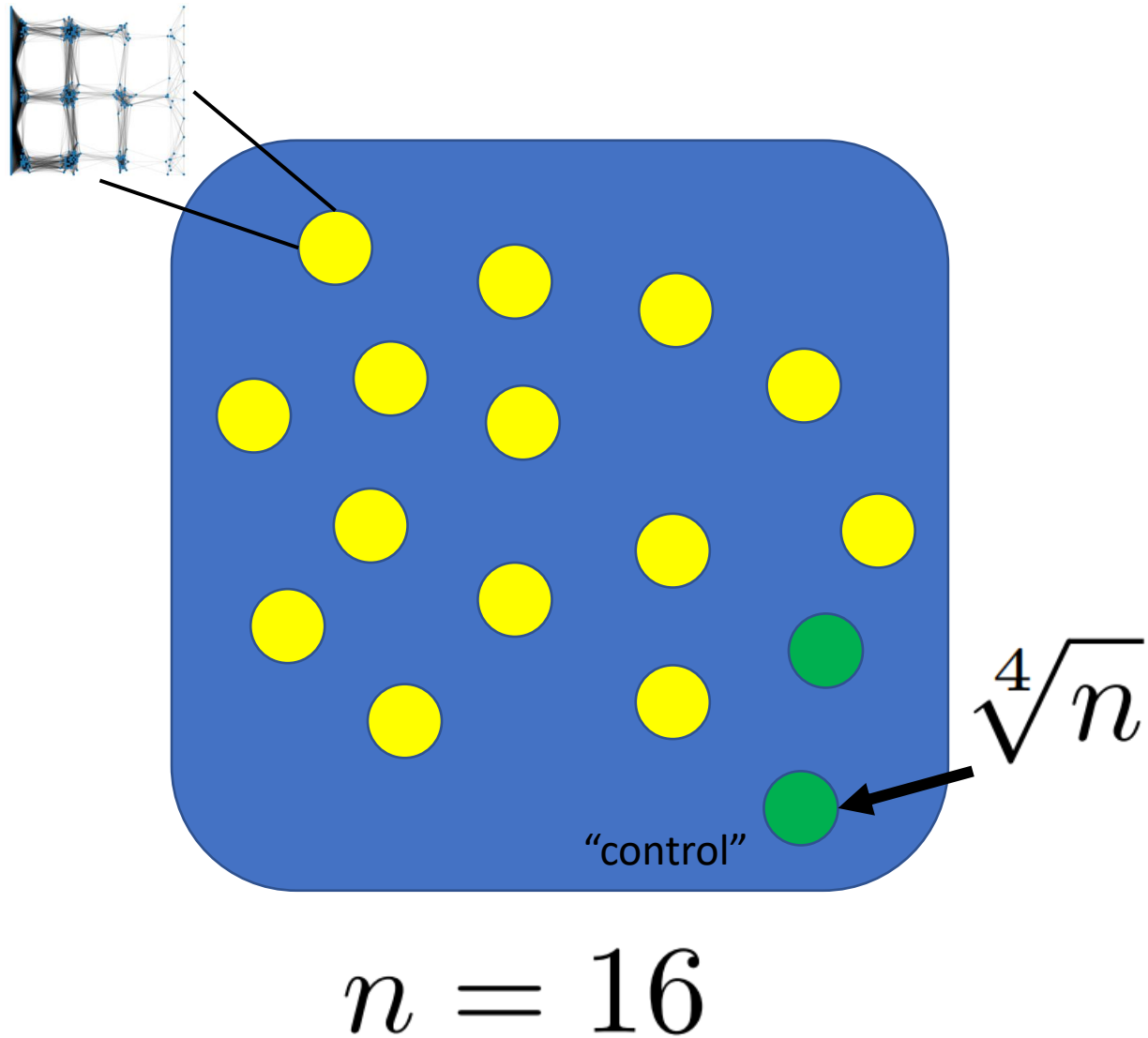
Applications :  
Base ImageNet / Pascal2  
Prédiction  
Detection d'objet

# *Functionalizing the Network using Evolution (Pseudo-Darwinism)*

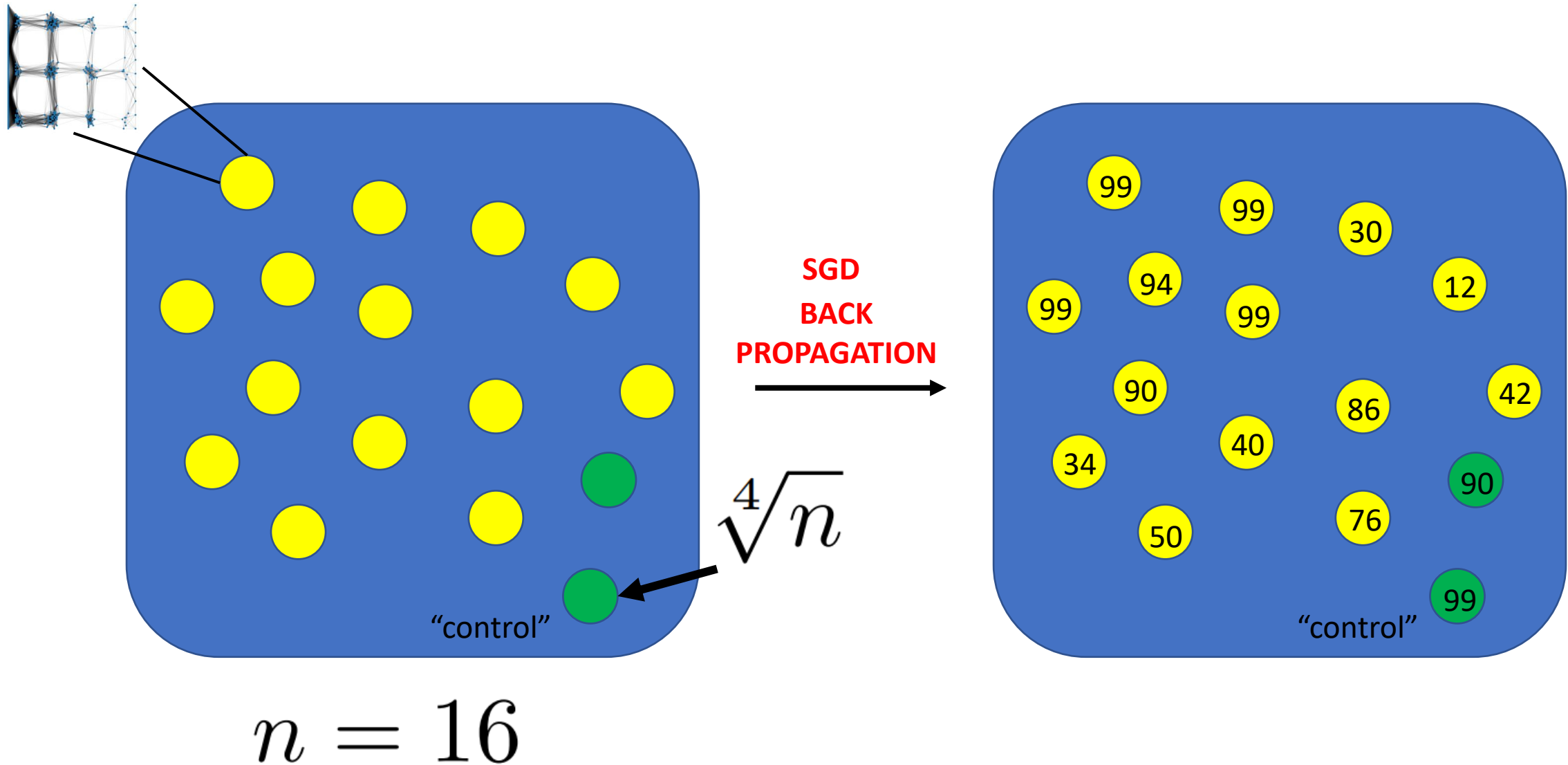


$$n = 16$$

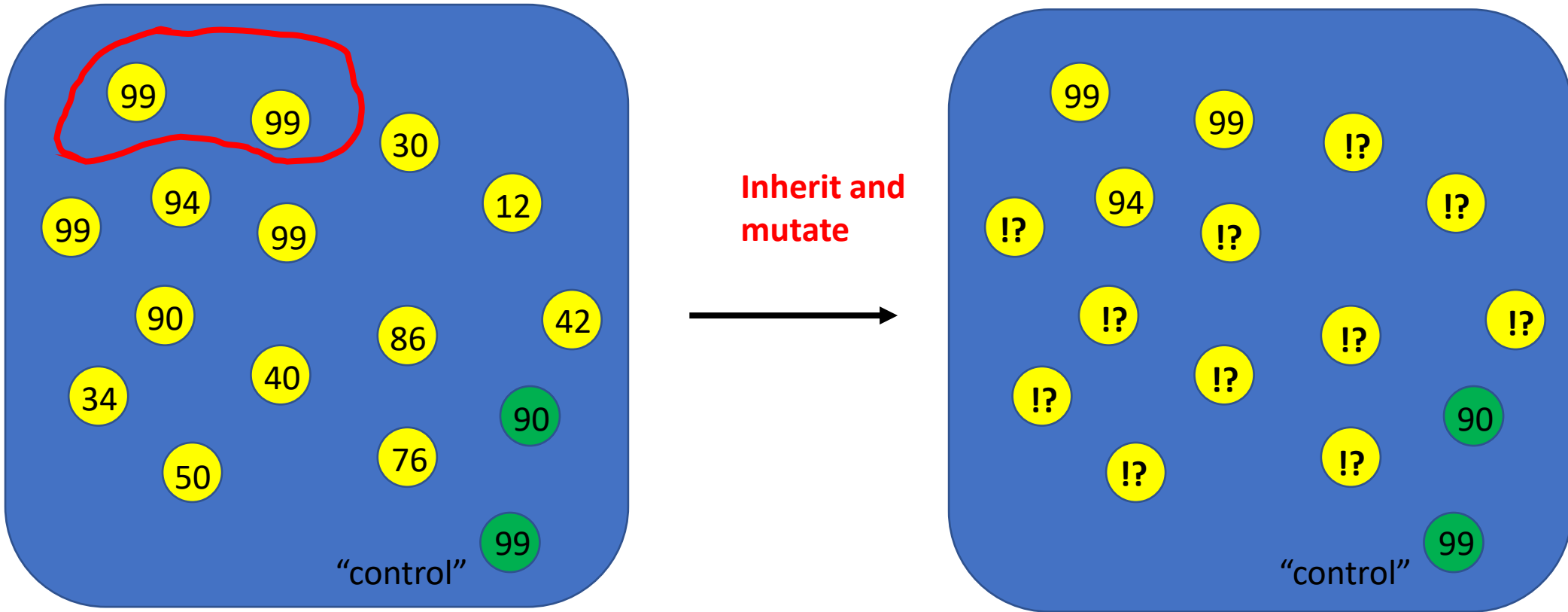
# Functionalizing the Network using Evolution (Pseudo-Darwinism)



# Functionalizing the Network using Evolution (Pseudo-Darwinism)



## Functionalizing the Network using Evolution (Pseudo-Darwinism)



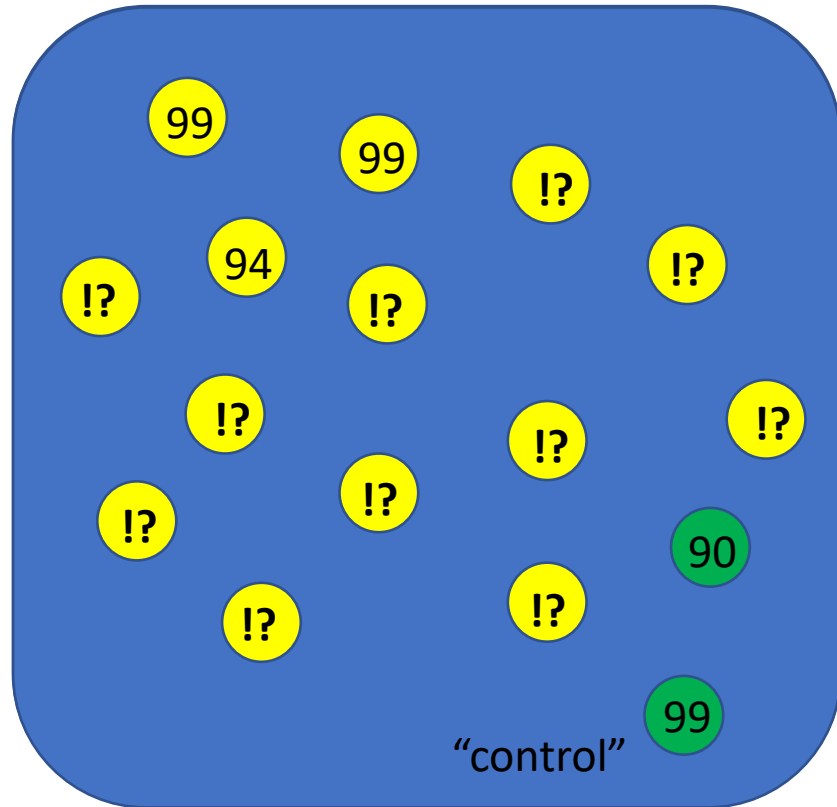
# CROSSOVER AND INHERIT TRAITS



## MUTATE



# Functionalizing the Network using Evolution (Pseudo-Darwinism)



3 Parents from G - 1

9 Offspring (inherited + mutated)

2 New Networks (random)

2 Control

$$\alpha = \sqrt{n} - 1$$

$$\alpha^2$$

$$\alpha$$

$$\sqrt[4]{n}$$

**CROSSOVER AND  
INHERIT TRAITS**

99

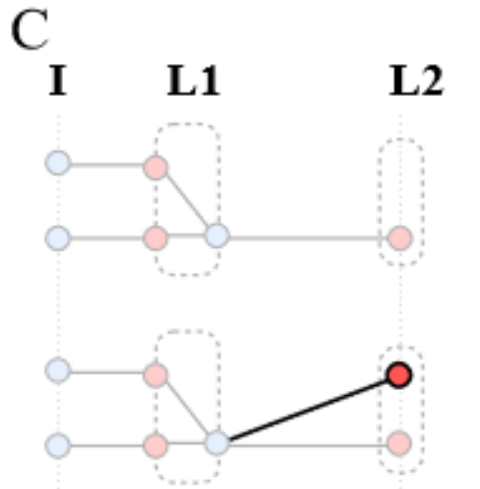
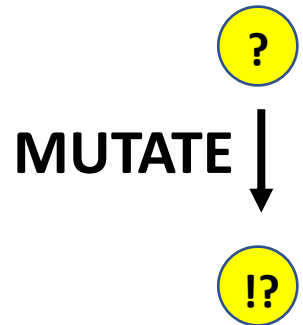
94

?

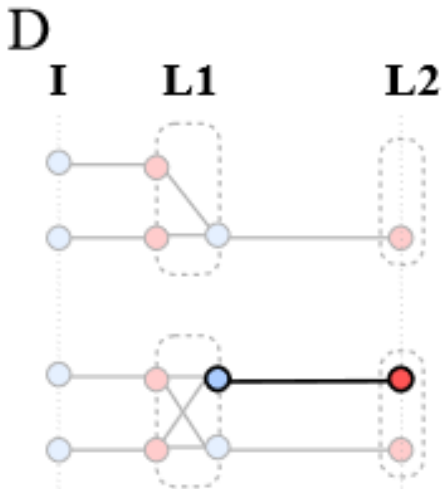
**MUTATE**

!?

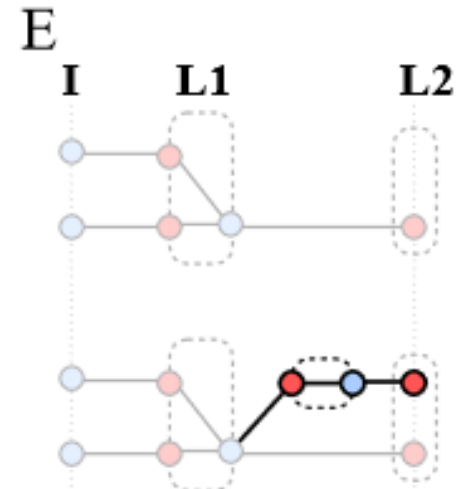
# All Possible Mutations to change network topology



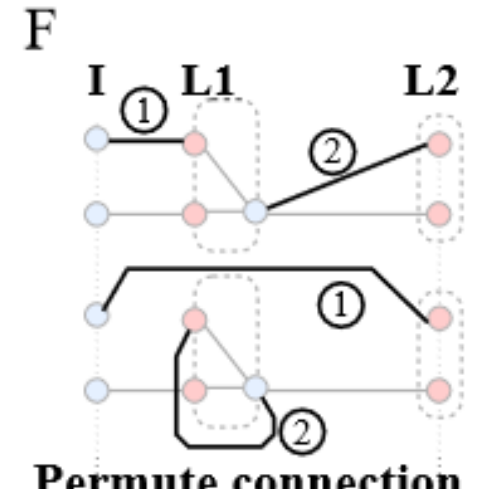
ADD/DEL connection



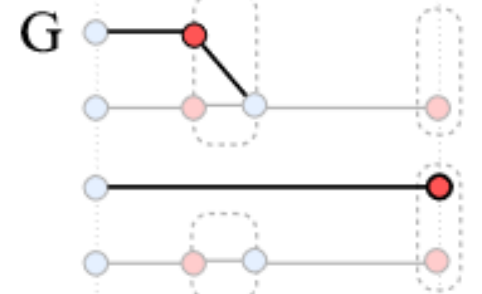
ADD/DEL neuron



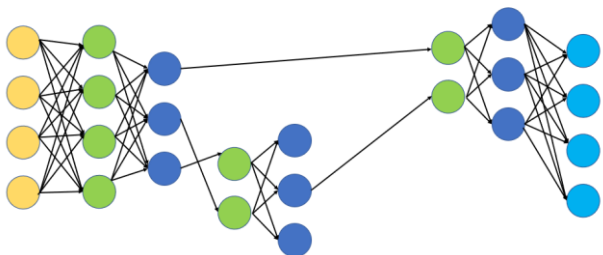
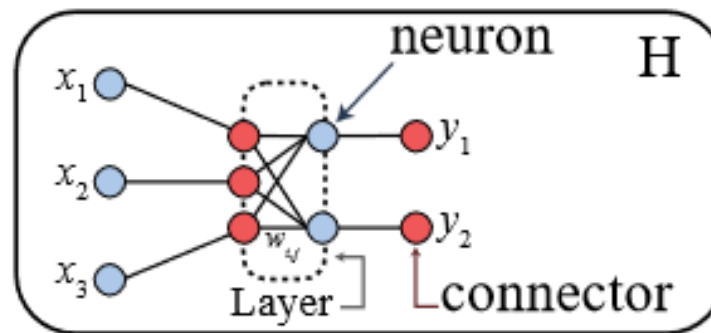
ADD Layer



Permute connection



Transposition



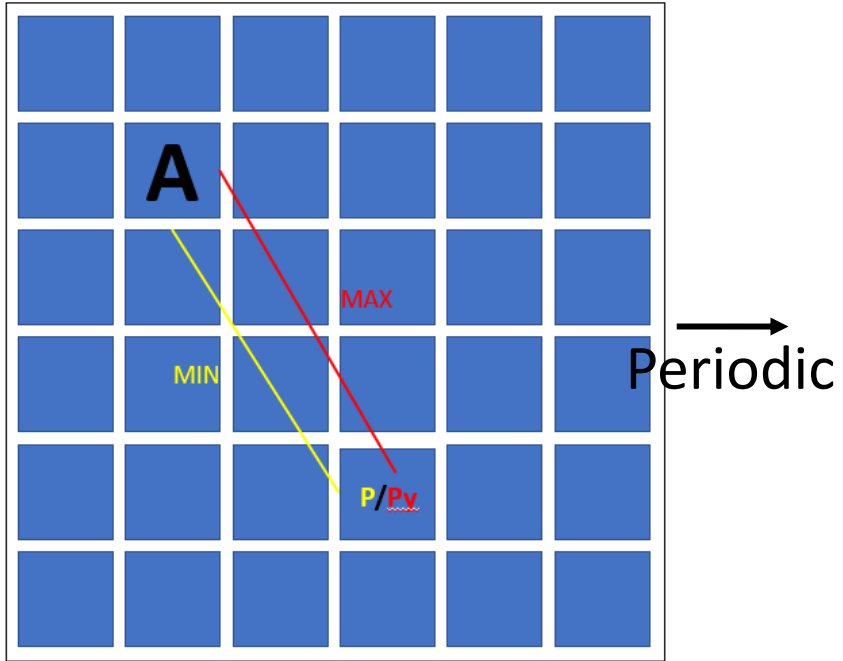


# Algorithm summary

1. draw  $n$  random realizations of neural networks, from which  $\lfloor \sqrt[4]{n} \rfloor$  will serve as control networks. This correspond to the initial population networks at the first generation ( $G_0$ )
2. train the current population using gradient descent and a number of batches
3. select the  $\alpha$  most efficient networks. The algorithm stops if either of the following two conditions is met: (1) the accuracy/score is 'good enough'; (2) the predefined maximal number of generations  $N$  is reached. This threshold number is defined by the number of batches required for the control networks for solving the same problem. Otherwise we continue.
4. create  $\alpha^2$  independent children, inheriting the structure (traits) from the selected parents (plasticity steps) networks and mutate their structures
5. insert  $\alpha$  new random networks
6. return to step (2) with this new population of  $n$  networks and start a new generation.

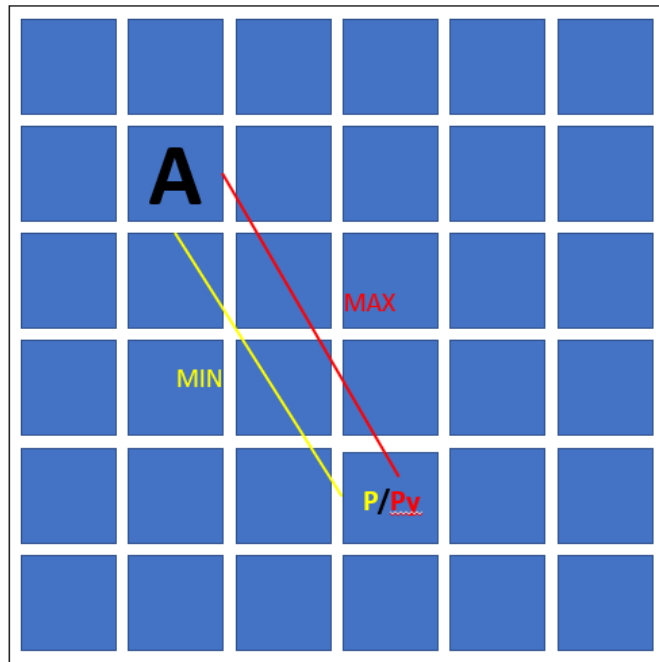


# *Tasks used to benchmark learning algorithm*



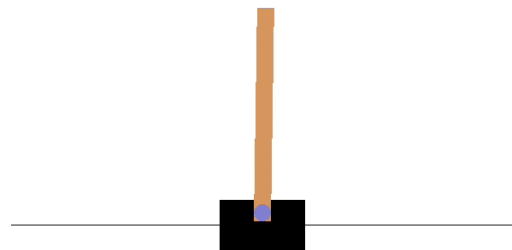
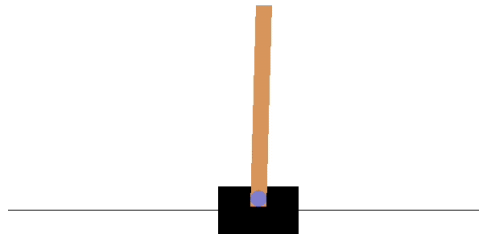
# Tasks used to benchmark learning algorithm

## TAG

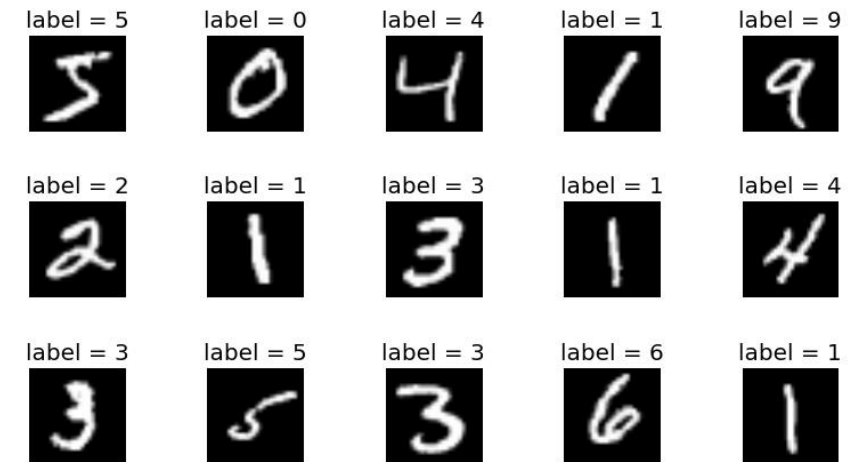


→  
Periodic

## CART-POLE



## MNIST



# *Tag-Game performance and functionalization*

12800 episodes

100 generations

128 episodes/gen

256 time  
steps/episode

1 batch (for SGD)  
is 16 time steps

relative overall stability

# Tag-Game performance and functionalization

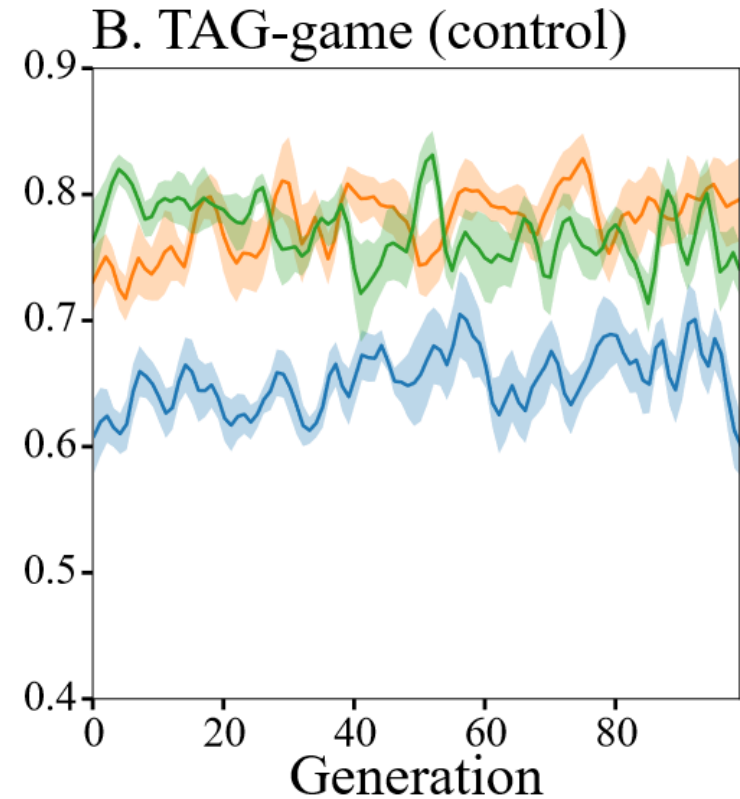
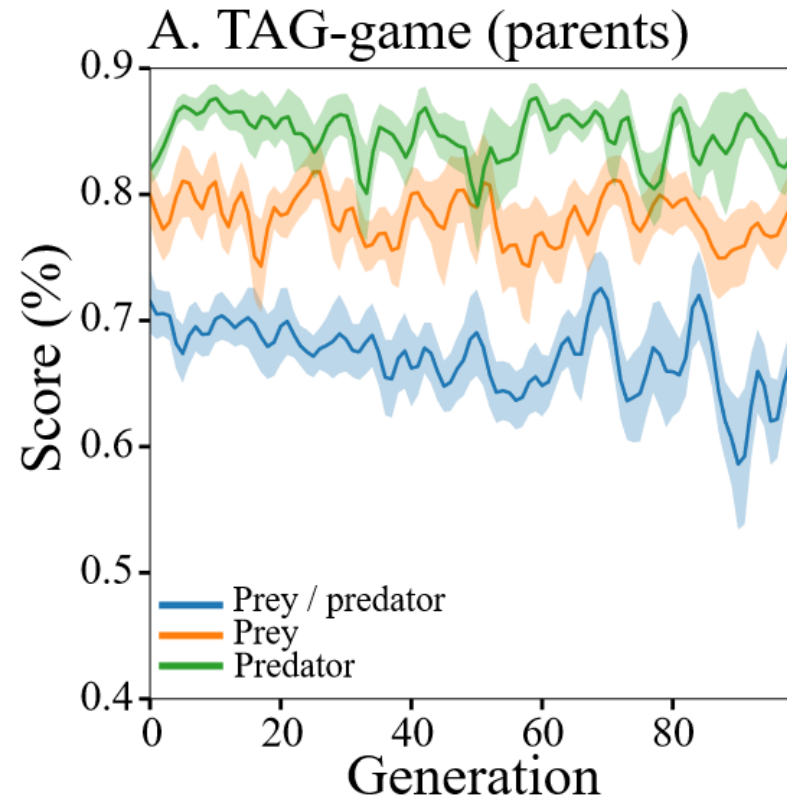
12800 episodes

100 generations

128 episodes/gen

256 time  
steps/episode

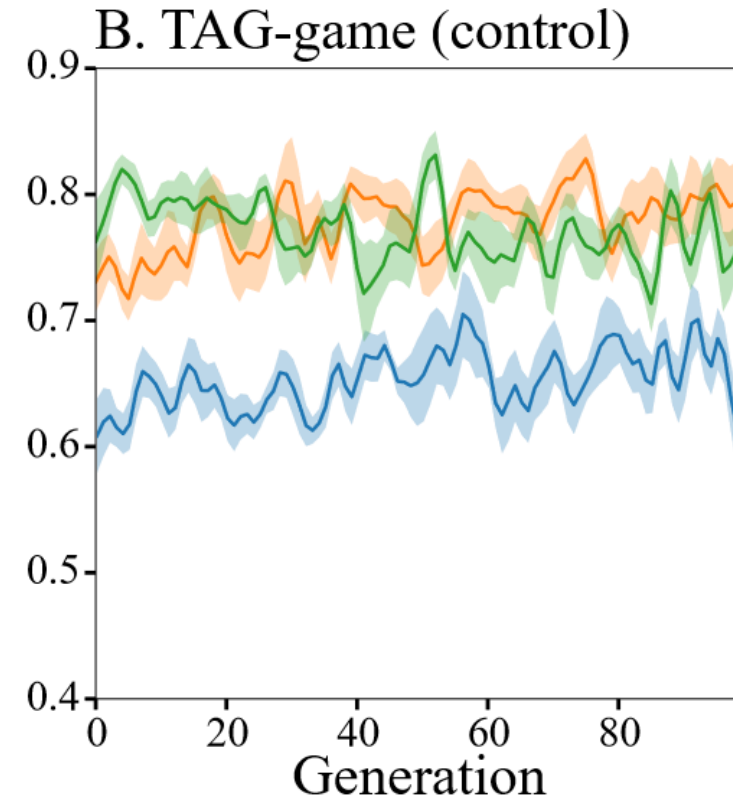
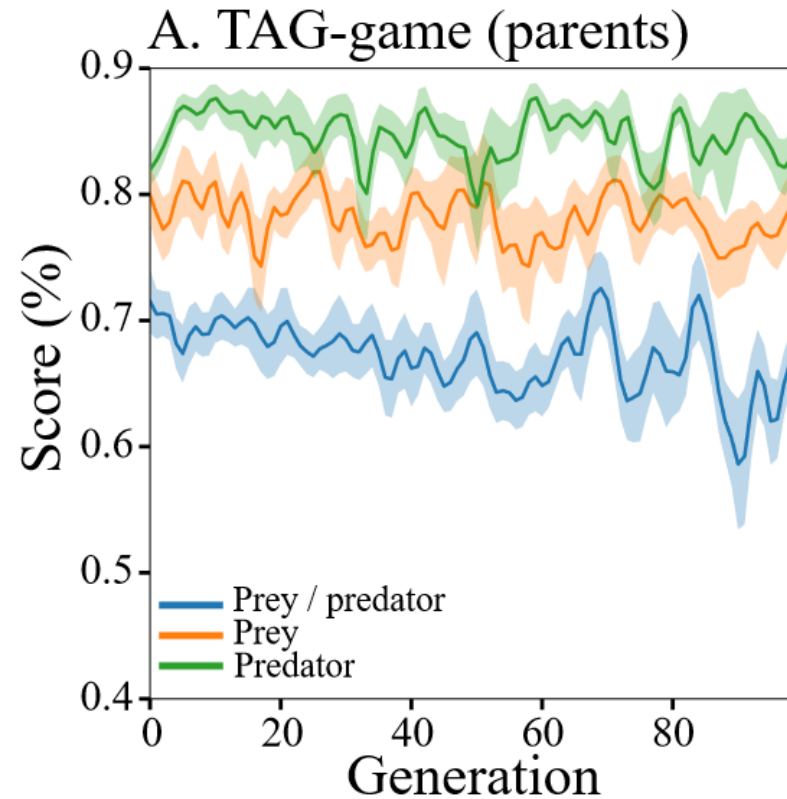
1 batch (for SGD)  
is 16 time steps



relative overall stability

# Tag-Game performance and functionalization

12800 episodes  
100 generations  
128 episodes/gen  
256 time  
steps/episode  
1 batch (for SGD)  
is 16 time steps



*'it appears that learning only one role -either be the prey or predator- is easier than learning both roles at once.'*

relative overall stability

# *Cart-Pole performance and functionalization*

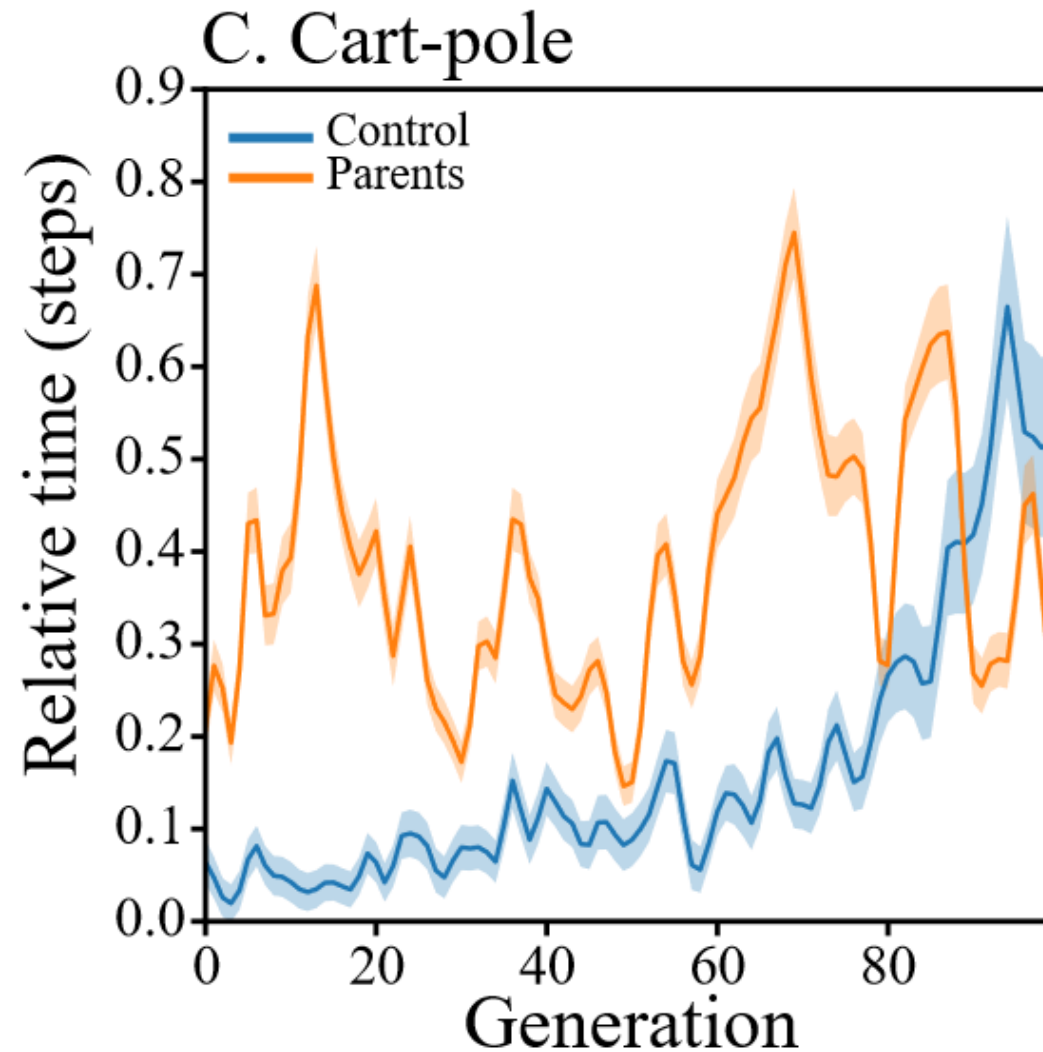
25000 episodes

100 generations

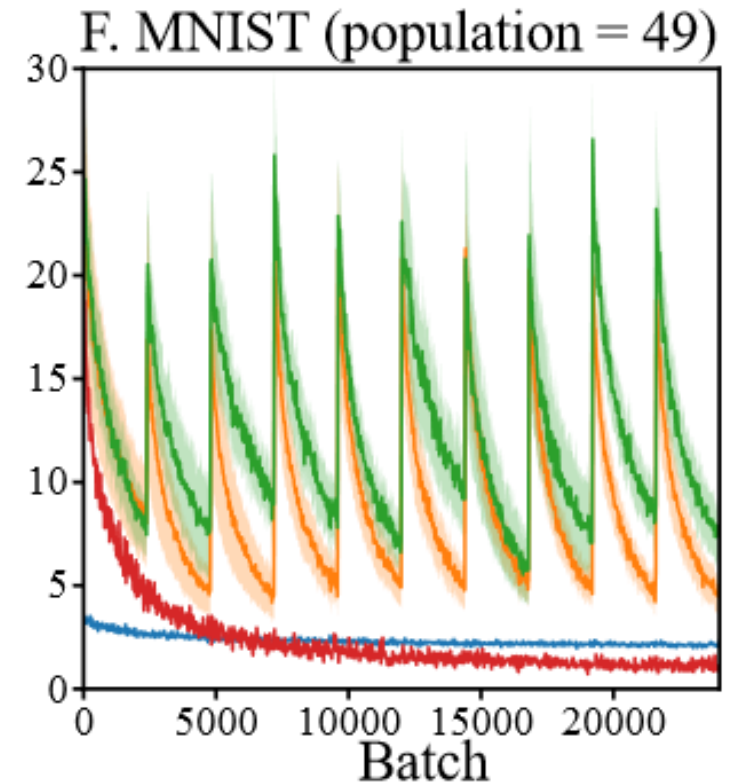
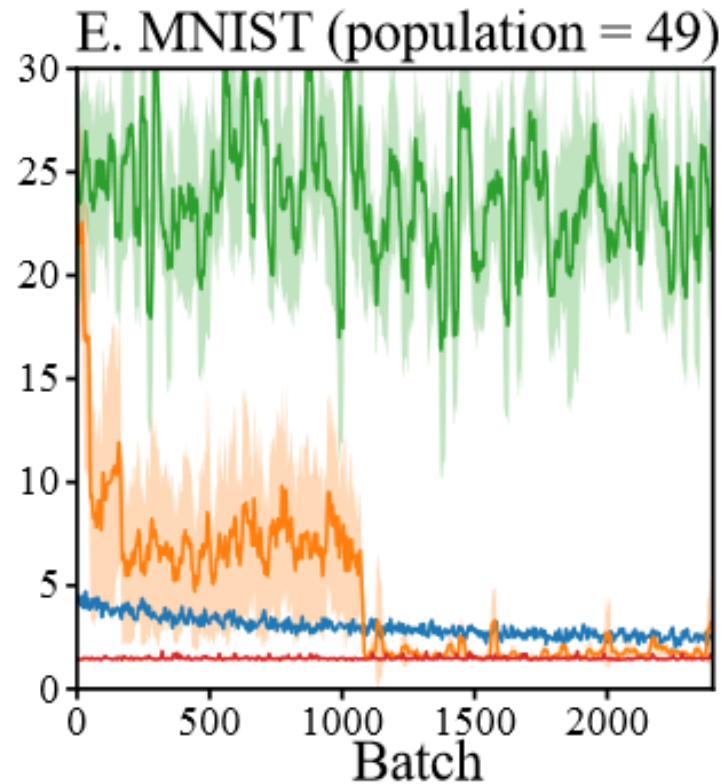
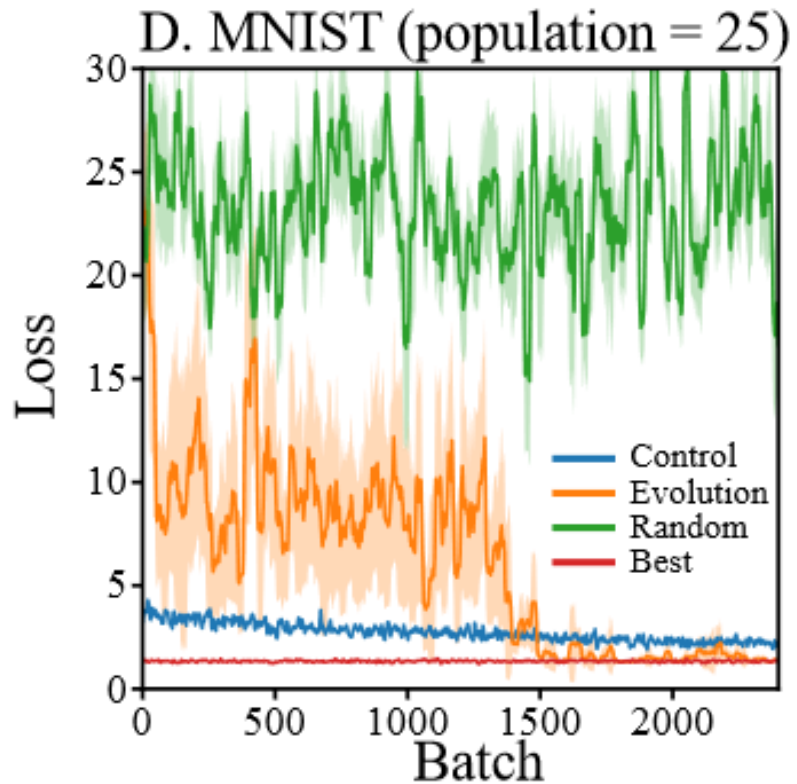
250 episodes/gen

300 time  
steps/episode

1 batch (for SGD)  
is 25 time steps

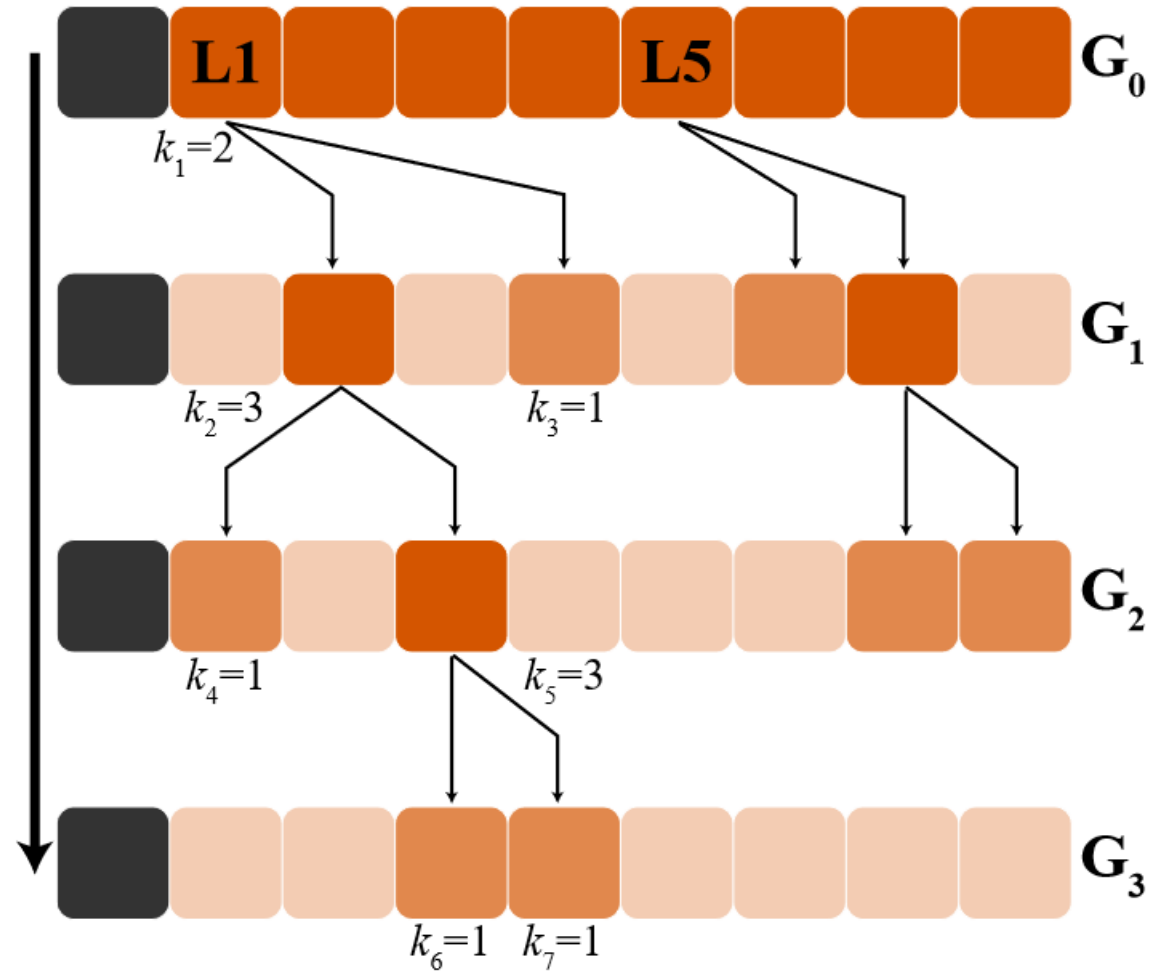
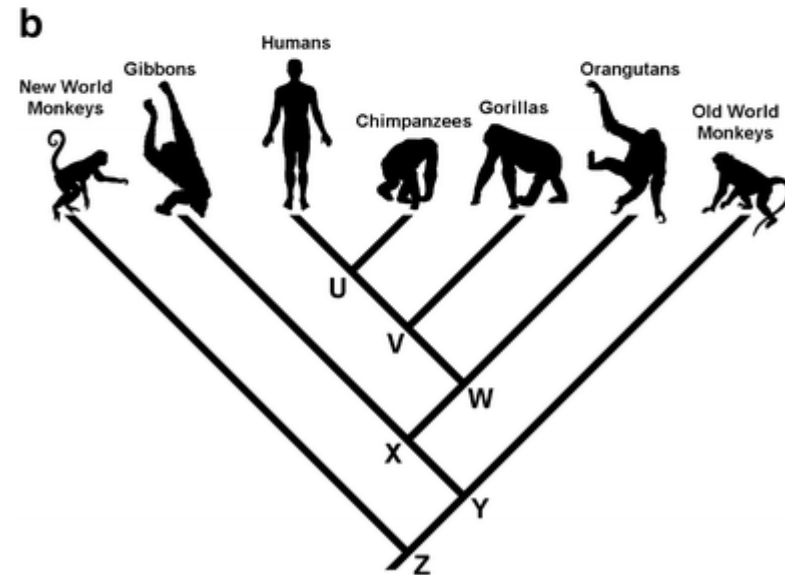
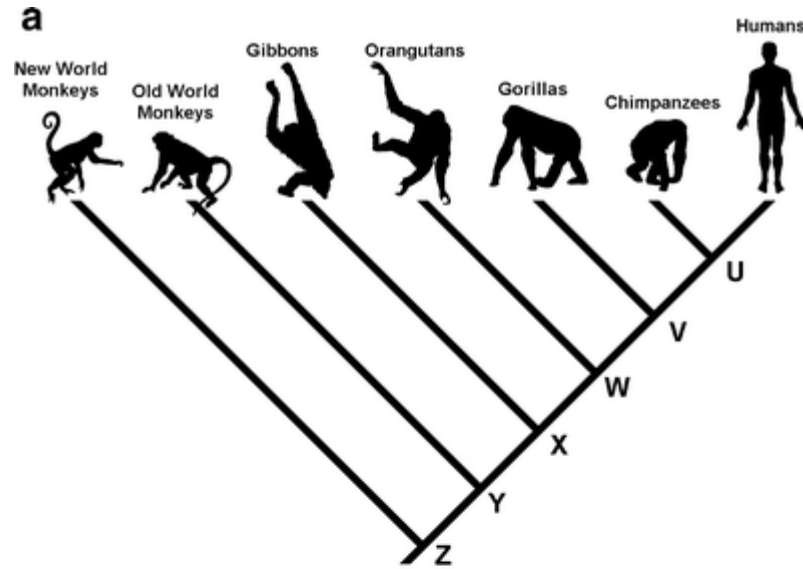


# *MNIST Classification loss vs batch-size*



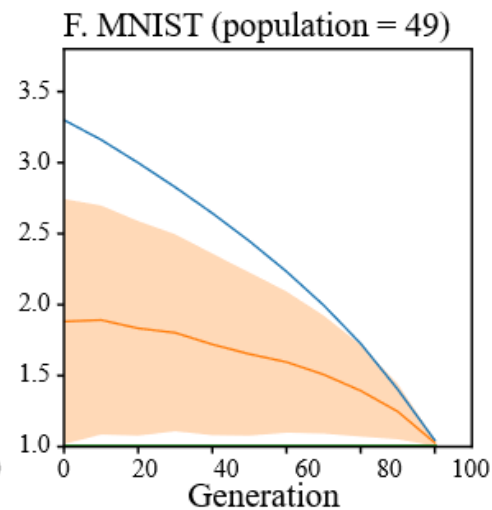
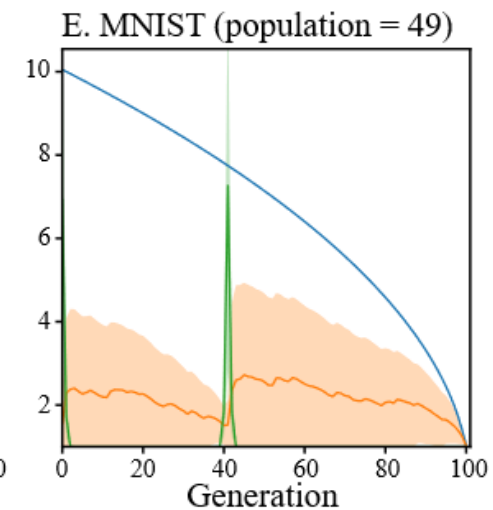
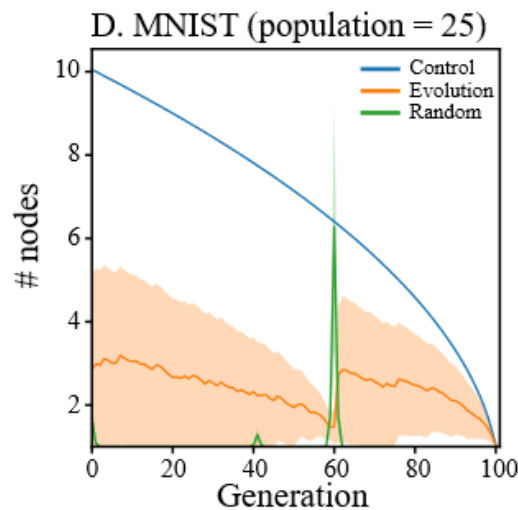
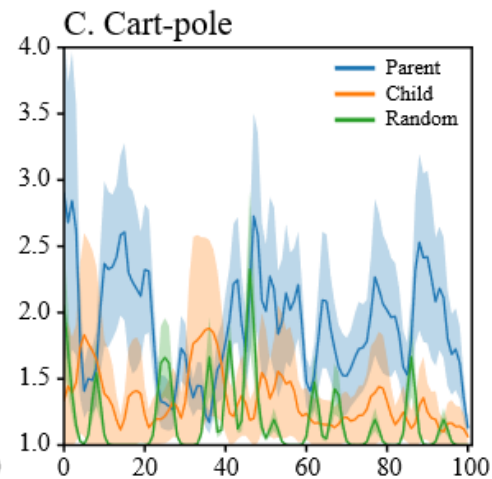
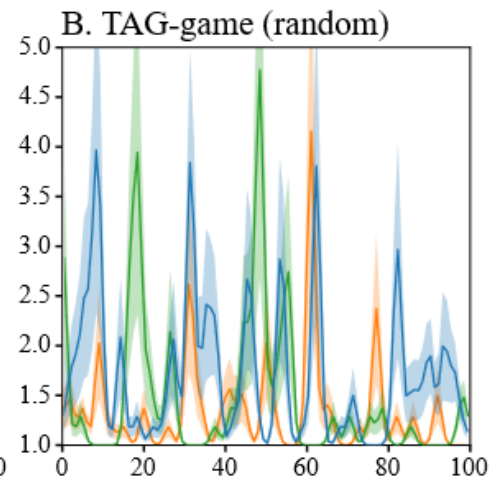
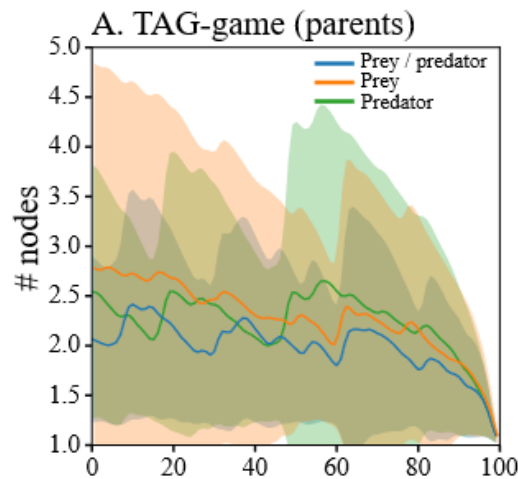
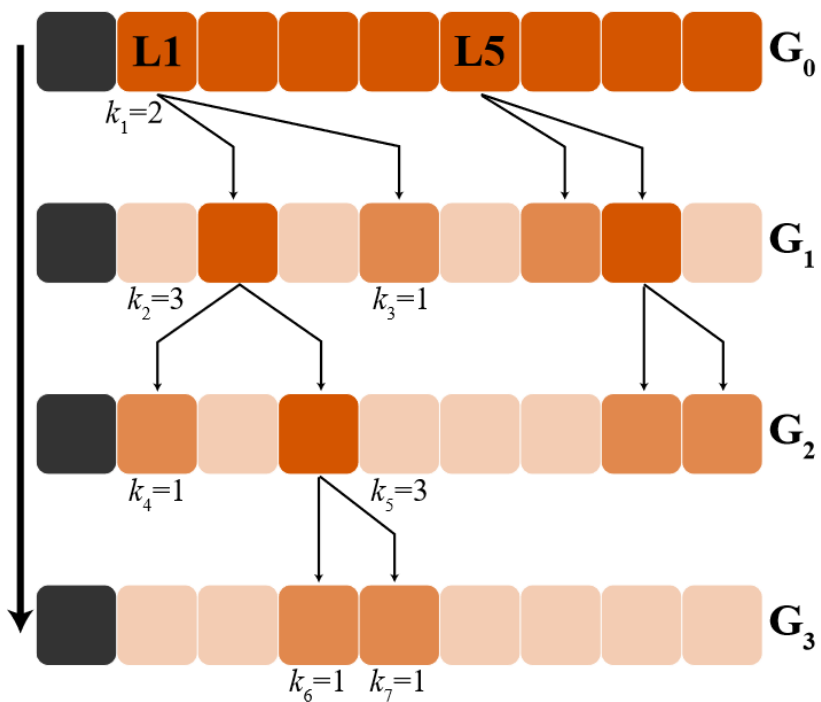
When this 'structural convergence' occurs, the evolutionary networks become more efficient than both the control and the random ones. The structural convergence occurs faster when we have a population of 49 networks (1100 batches) than when we have a population of 25 networks (1500 batches).

# Evolutionary Stability of a lineage in the population





# Evolutionary Stability of a lineage in the population

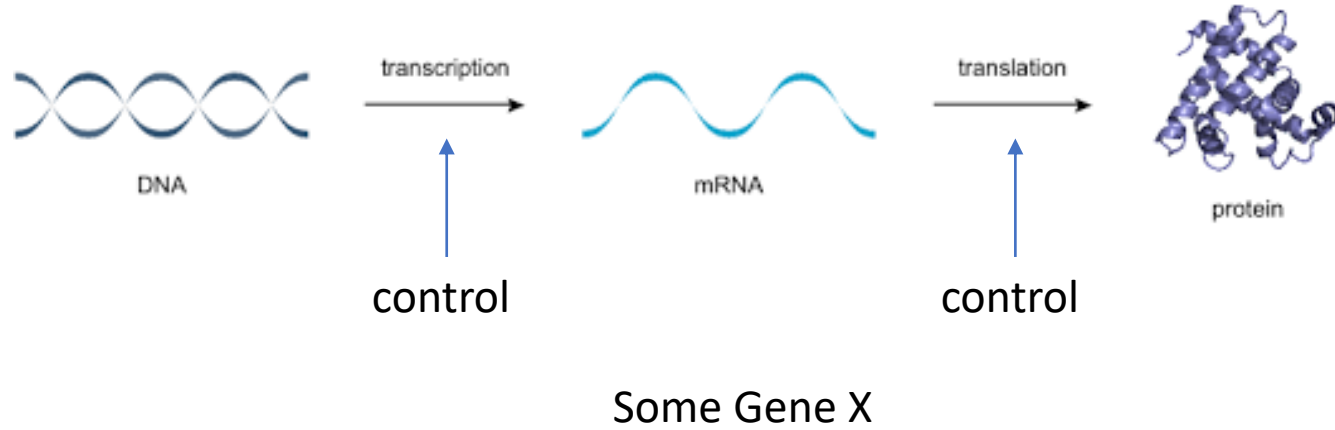


# Conclusions

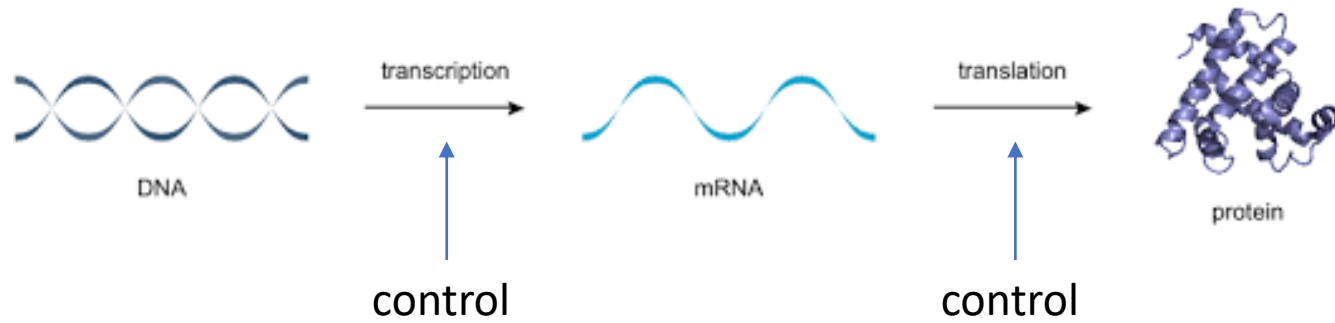
- (i) the larger the network, the more evolutionarily stable it is.
- (i) the greater the number of individuals in the initial population, the faster it converges.
- (ii) A greater efficiency when separating roles in the TAG-game, suggesting that specialization is more effective
- (iii) Two types of convergence can be obtained: the first one is the 'structural convergence' occurring when increasing the number of mutations (or generations). The second one can be obtained by gradient descent when there is less mutations and more batches

Can Network Stability promise  
Evolutionary Stability?

# *Robust Motifs are by Evolutionary Design*



# *Robust Motifs are by Evolutionary Design*

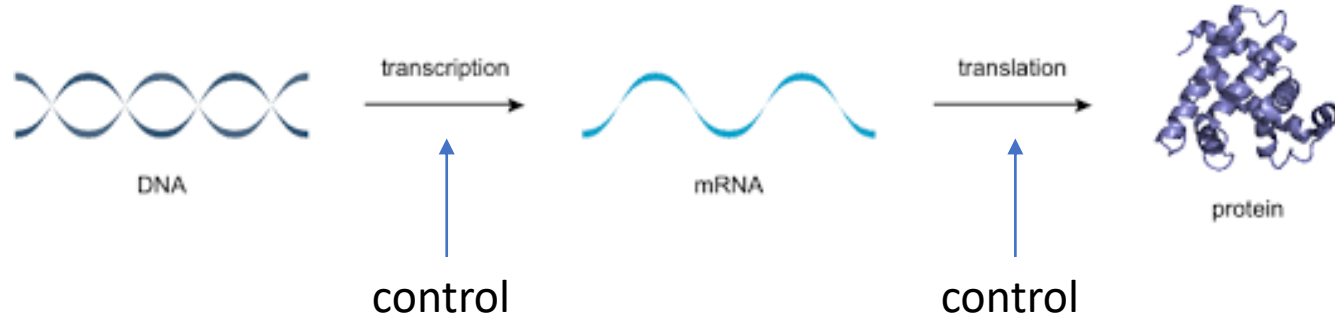


Some Gene X

Coherent  
feed-forward loop  
(C1-FFL)

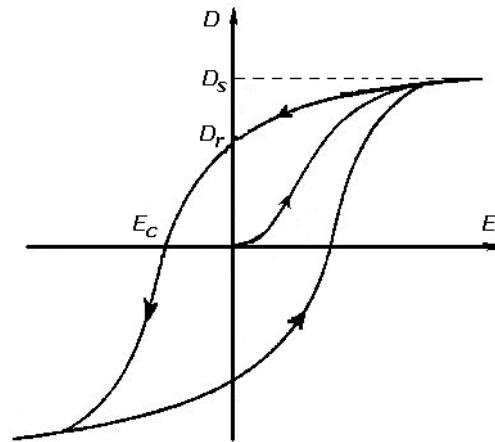


# Robust Motifs are by Evolutionary Design



Some Gene X

Coherent  
feed-forward loop  
(C1-FFL)



(Hysteresis)

# Robust Motifs are by Evolutionary Design

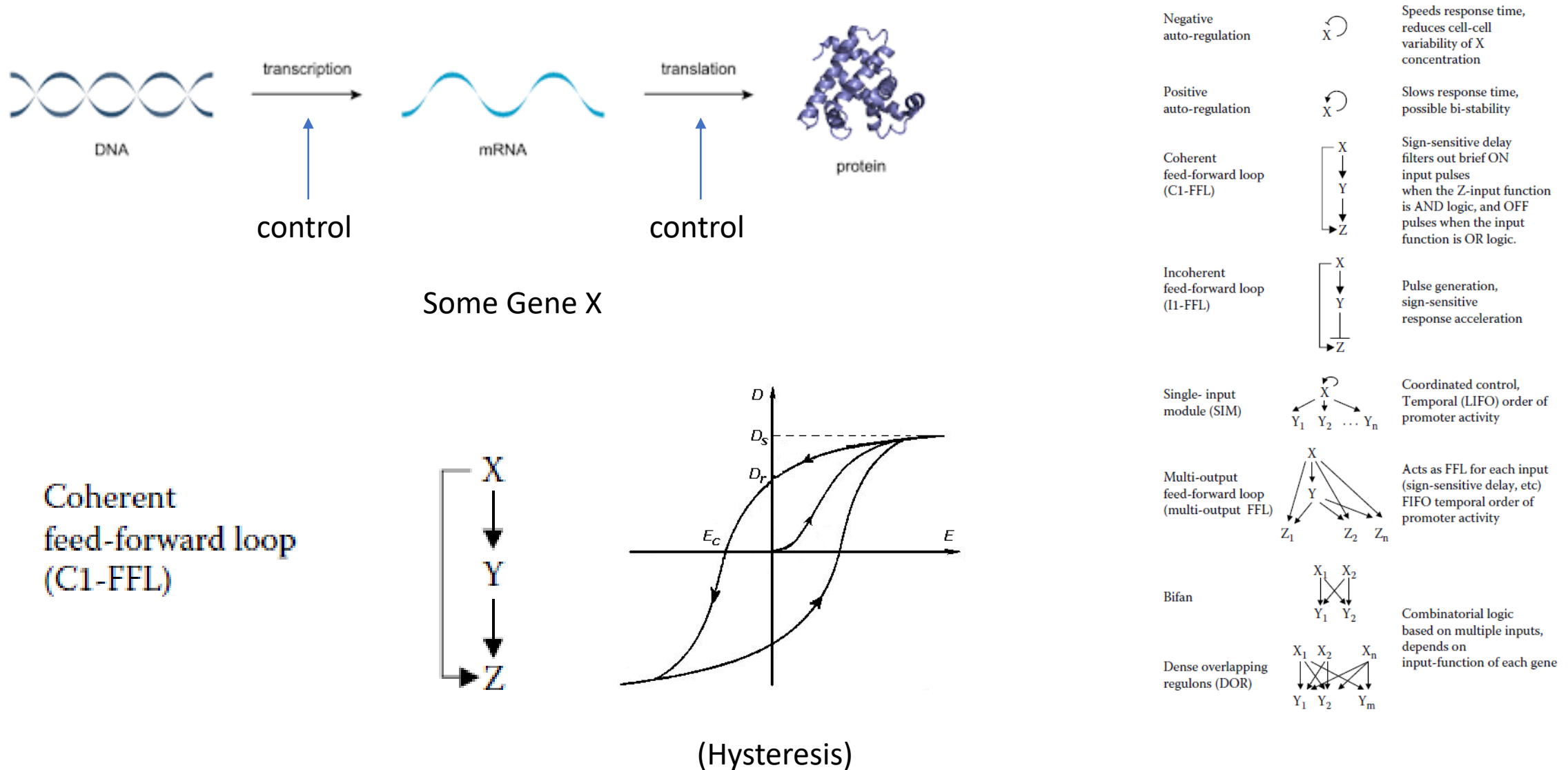
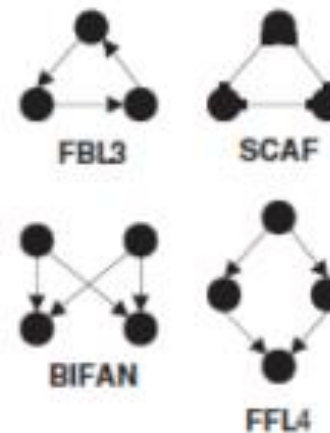
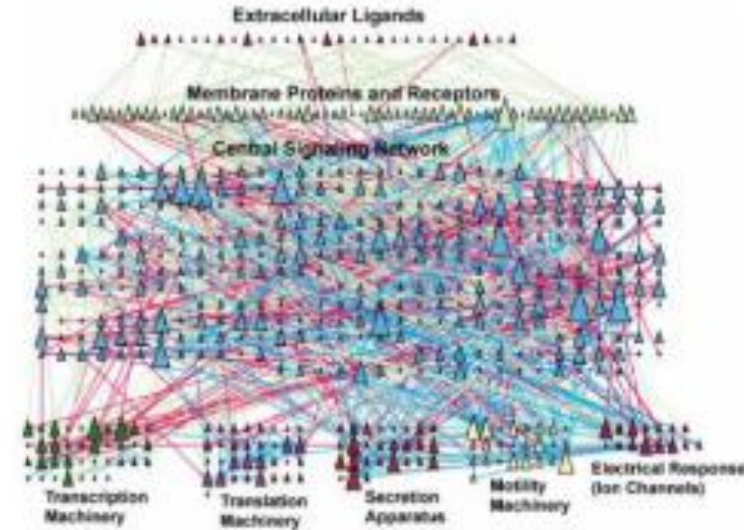
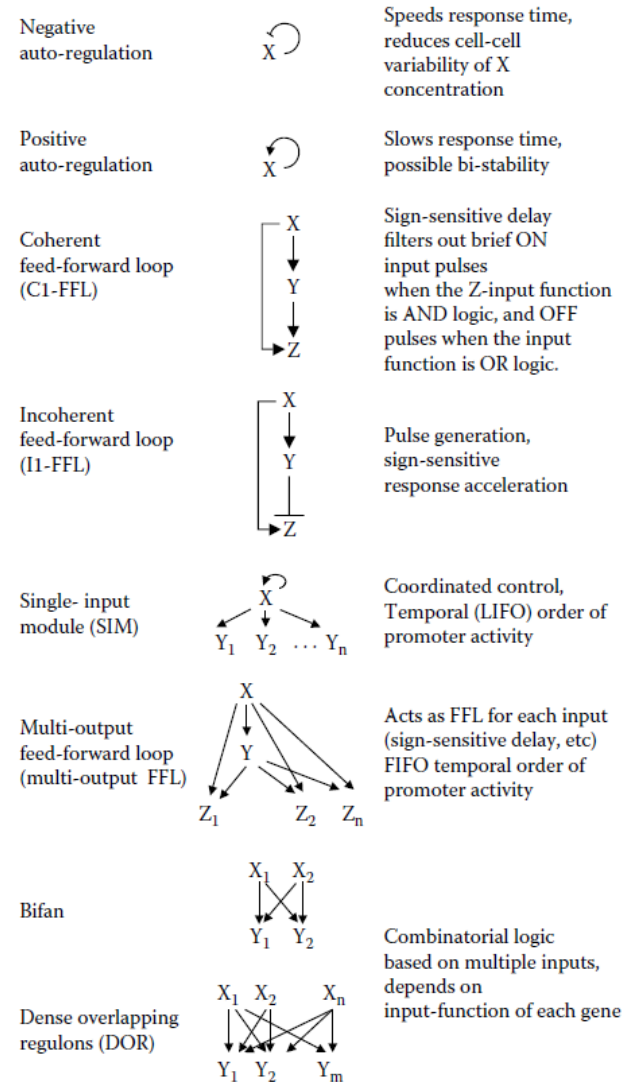


FIGURE 5.15 Network motifs in sensory transcription networks.

# Robust Motifs are by Evolutionary Design



Motif #	Motifs counts		
	CN*	SN**	Z-score
FBL3	22	8.6 ± 8.1	4.34
SCAF	25	9.0 ± 3.4	4.72
BIFAN	1009	180.3 ± 28.0	29.60
FFL4	301	103.4 ± 17.2	11.51

\* CN- Cellular Network.  
 \*\* SN- Shuffled Networks.  
 Mean ± SD computed for 100 shuffled networks.

FIGURE 5.15 Network motifs in sensory transcription networks.

Ma'ayan A et al (2005) Science 309: 1078



# Information Flow Stability informed Evolutionary Learning

The diagram illustrates the relationship between a neural network structure, a population of nodes, and a weight matrix  $W$ .

On the left, a small inset shows a 3D grid of nodes connected by edges. Below it, a large blue rounded rectangle contains a population of nodes. Most nodes are yellow, but two nodes at the bottom right are green. One green node is labeled "control" with an arrow pointing to it. A line connects one of the yellow nodes in the population to a specific row and column in the weight matrix  $W$ .

On the right, the weight matrix  $W$  is shown as a table with rows and columns labeled B, I, L1, L2, and O. The columns are also labeled I, L1, L2, and O. The matrix contains binary values (0 or 1) and some cells are highlighted in yellow. A red dashed diagonal line runs from the top-left to the bottom-right.

		I	L1	L2	O
B		0	0	0	0
I		0	0	0	0
L1		0	0	0	0
L2		0	0	0	0
O		0	0	0	0

The matrix  $W$  is defined by the following values (rows are labeled on the left, columns on top):

- Row B: [0, 0, 0, 0]
- Row I: [0, 0, 0, 0]
- Row L1: [0, 0, 0, 0]
- Row L2: [0, 0, 0, 0]
- Row O: [0, 0, 0, 0]

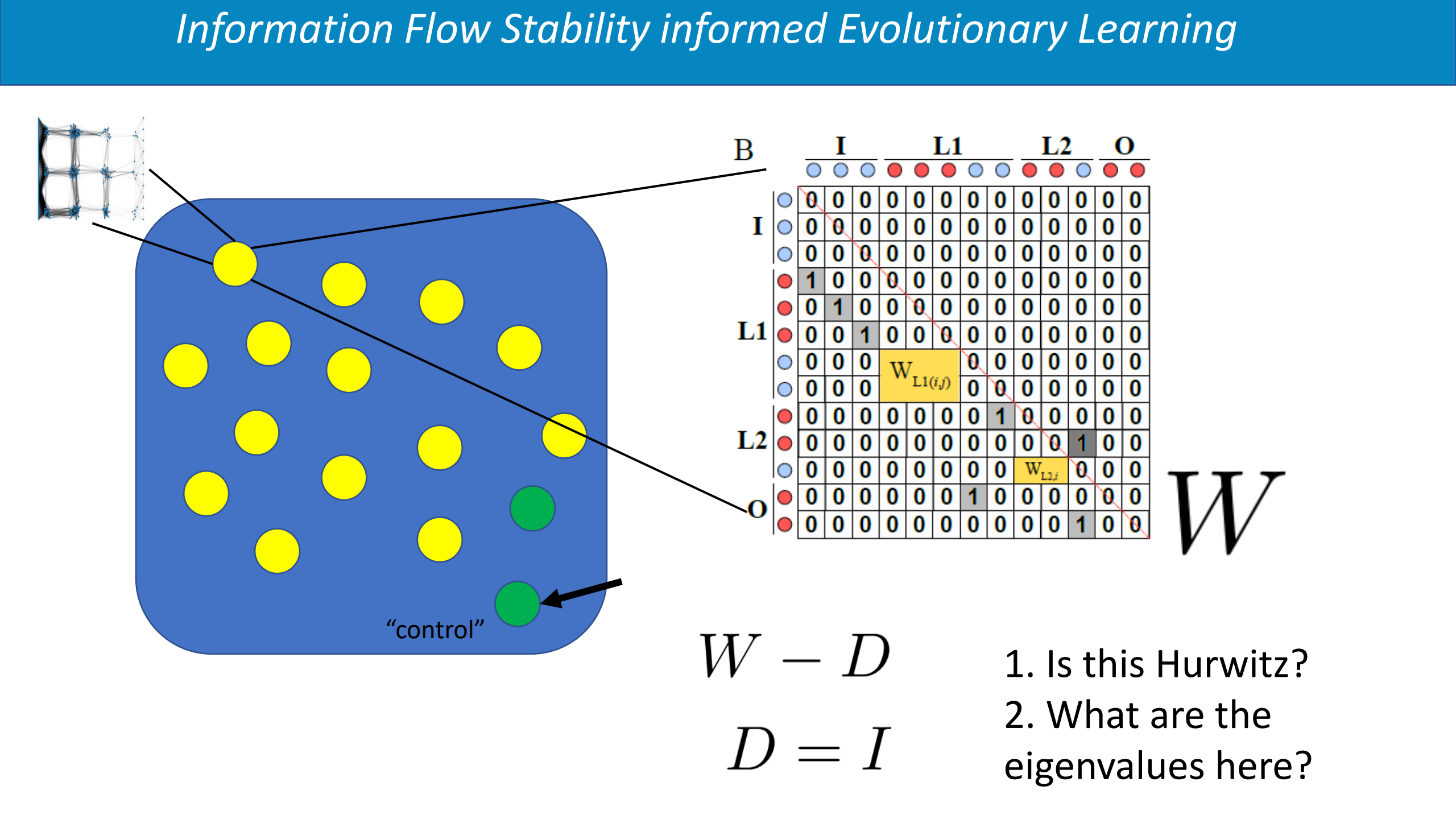
Specific cells in the matrix are highlighted in yellow:

- Row L1, Column I:  $W_{L1(i,j)}$
- Row L2, Column L2:  $W_{L2,j}$

$$W - D$$

$$D = I$$

1. Is this Hurwitz?
2. What are the eigenvalues here?



# Information Flow Stability informed Evolutionary Learning

The diagram illustrates the relationship between a neural network structure, a population of nodes, and a weight matrix  $W$ .

On the left, a small inset shows a 3D grid of nodes connected by edges. Below it, a large blue rounded rectangle contains a population of nodes. Most nodes are yellow, but two nodes at the bottom right are green. An arrow points to one of these green nodes with the label "control".

On the right, a weight matrix  $W$  is shown. The matrix is a 12x12 grid with columns labeled I, L1, L2, and O. The rows are labeled B, I, L1, L2, and O. The matrix contains 0s and 1s, with some cells highlighted in yellow. The yellow cells are labeled  $W_{L1(i,j)}$  and  $W_{L2,j}$ . A red dashed line runs diagonally from the top-left to the bottom-right of the matrix.

$$W - D$$

$$D = I$$

1. Is this Hurwitz?
2. What are the eigenvalues here?

# Information Flow Stability informed Evolutionary Learning

The diagram illustrates the relationship between a neural network structure, a population of nodes, and a weight matrix  $W$ .

On the left, a small inset shows a 3D grid of nodes connected by edges. Below it, a large blue rounded rectangle contains a population of nodes. Most nodes are yellow, but two nodes at the bottom right are green. One green node is labeled "control" with an arrow pointing to it. A line connects one of the yellow nodes in the population to a specific row and column in the weight matrix  $W$ .

On the right, the weight matrix  $W$  is shown as a table with rows and columns labeled B, I, L1, L2, and O. The columns are also labeled I, L1, L2, and O. The matrix contains binary values (0 or 1) and some cells are highlighted in yellow or gray. A red dashed line runs diagonally from the top-left to the bottom-right of the matrix.

		I	L1	L2	O
B		0	0	0	0
I		0	0	0	0
L1		0	0	0	0
L2		0	0	0	0
O		0	0	0	0

The matrix  $W$  is defined by the following values (rows are labeled on the left, columns on top):

- Row B: [0, 0, 0, 0]
- Row I: [0, 0, 0, 0]
- Row L1: [0, 0, 0, 0]
- Row L2: [0, 0, 0, 0]
- Row O: [0, 0, 0, 0]

Specific cells in the matrix are highlighted:

- Row I, Column I: 1 (gray)
- Row L1, Column L1: 1 (gray)
- Row L2, Column L2: 1 (gray)
- Row O, Column O: 1 (gray)

Yellow cells in the matrix indicate weights  $W_{L1(i,j)}$  and  $W_{L2,j}$ .

$$W - D$$

$$D = I$$

1. Is this Hurwitz?
2. What are the eigenvalues here?

# Information Flow Stability informed Evolutionary Learning

The diagram illustrates the relationship between a neural network structure, a population of nodes, and a weight matrix  $W$ .

On the left, a small inset shows a 3D grid of nodes connected by edges. Below it, a large blue rounded rectangle contains a population of nodes. Most nodes are yellow, but two nodes at the bottom right are green. One green node is labeled "control" with an arrow pointing to it. A line connects one of the yellow nodes in the population to a specific row and column in the weight matrix  $W$ .

On the right, the weight matrix  $W$  is shown as a table with rows and columns labeled B, I, L1, L2, and O. The columns are also labeled I, L1, L2, and O. The matrix contains binary values (0 or 1) and some cells are highlighted in yellow or gray. A red dashed line runs diagonally from the top-left to the bottom-right of the matrix.

		I	L1	L2	O
B		0	0	0	0
I		0	0	0	0
L1		0	0	0	0
L2		0	0	0	0
O		0	0	0	0

The matrix  $W$  is defined by the following values (rows are labeled on the left, columns on top):

- Row B: [0, 0, 0, 0]
- Row I: [0, 0, 0, 0]
- Row L1: [0, 0, 0, 0]
- Row L2: [0, 0, 0, 0]
- Row O: [0, 0, 0, 0]

Specific cells in the matrix are highlighted:

- Row I, Column I: 1 (gray)
- Row L1, Column L1: 1 (gray)
- Row L2, Column L2: 1 (gray)
- Row O, Column O: 1 (gray)

Yellow cells in the matrix indicate weights  $W_{L1(i,j)}$  and  $W_{L2,j}$ .

$$W - D$$

$$D = I$$

1. Is this Hurwitz?
2. What are the eigenvalues here?

# Information Flow Stability informed Evolutionary Learning

The diagram illustrates the relationship between a neural network structure, a population of nodes, and a weight matrix  $W$ .

On the left, a small inset shows a 3D grid of nodes connected by edges. Below it, a large blue rounded rectangle contains a population of nodes. Most nodes are yellow, but two nodes at the bottom right are green. One green node is labeled "control" with an arrow pointing to it. A line connects one of the yellow nodes in the population to a specific row and column in the weight matrix  $W$ .

On the right, the weight matrix  $W$  is shown as a table with rows and columns labeled B, I, L1, L2, and O. The columns are also labeled I, L1, L2, and O. The matrix contains binary values (0 or 1) and some cells are highlighted in yellow or gray. A red dashed line runs diagonally from the top-left to the bottom-right of the matrix.

		I	L1	L2	O
B		0	0	0	0
I		0	0	0	0
L1		0	0	0	0
L2		0	0	0	0
O		0	0	0	0

The matrix  $W$  is defined by the following values (rows are labeled on the left, columns on top):

- Row B: [0, 0, 0, 0]
- Row I: [0, 0, 0, 0]
- Row L1: [0, 0, 0, 0]
- Row L2: [0, 0, 0, 0]
- Row O: [0, 0, 0, 0]

Specific cells in the matrix are highlighted:

- Row I, Column I: 1 (gray)
- Row L1, Column L1: 1 (gray)
- Row L2, Column L2: 1 (gray)
- Row O, Column O: 1 (gray)

The matrix  $W$  is labeled with a large  $W$  to its right.

$$W - D$$

$$D = I$$

1. Is this Hurwitz?
2. What are the eigenvalues here?

- # Information Flow Stability informed Evolutionary Learning
- The diagram illustrates the relationship between a neural network structure, a population of nodes, and a weight matrix  $W$ .

On the left, a small inset shows a 3D grid of nodes connected by edges. Below it, a large blue rounded rectangle contains a population of nodes. Most nodes are yellow, but two nodes at the bottom right are green. One of these green nodes is labeled "control" with an arrow pointing to it. A line connects one of the yellow nodes in the population to a specific row and column in the weight matrix  $W$ .

On the right, the weight matrix  $W$  is shown as a table with rows and columns labeled B, I, L1, L2, and O. The columns are also labeled I, L1, L2, and O. The matrix contains binary values (0 or 1) and some cells are highlighted in yellow or gray. A red dashed diagonal line runs from the top-left to the bottom-right of the matrix.

		I	L1	L2	O
B		0	0	0	0
I		0	0	0	0
L1		0	0	0	0
L2		0	0	0	0
O		0	0	0	0

The matrix  $W$  is defined by the following values (rows are labeled on the left, columns on top):

  - Row B: [0, 0, 0, 0]
  - Row I: [0, 0, 0, 0]
  - Row L1: [0, 0, 0, 0]
  - Row L2: [0, 0, 0, 0]
  - Row O: [0, 0, 0, 0]

Specific cells in the matrix are highlighted:

  - Row I, Column I: 1 (gray)
  - Row L1, Column L1: 1 (gray)
  - Row L2, Column L2: 1 (gray)
  - Row O, Column O: 1 (gray)

Yellow cells in the matrix indicate weights  $W_{L1(i,j)}$  and  $W_{L2,j}$ .
- $$W - D$$
- $$D = I$$
1. Is this Hurwitz?
  2. What are the eigenvalues here?

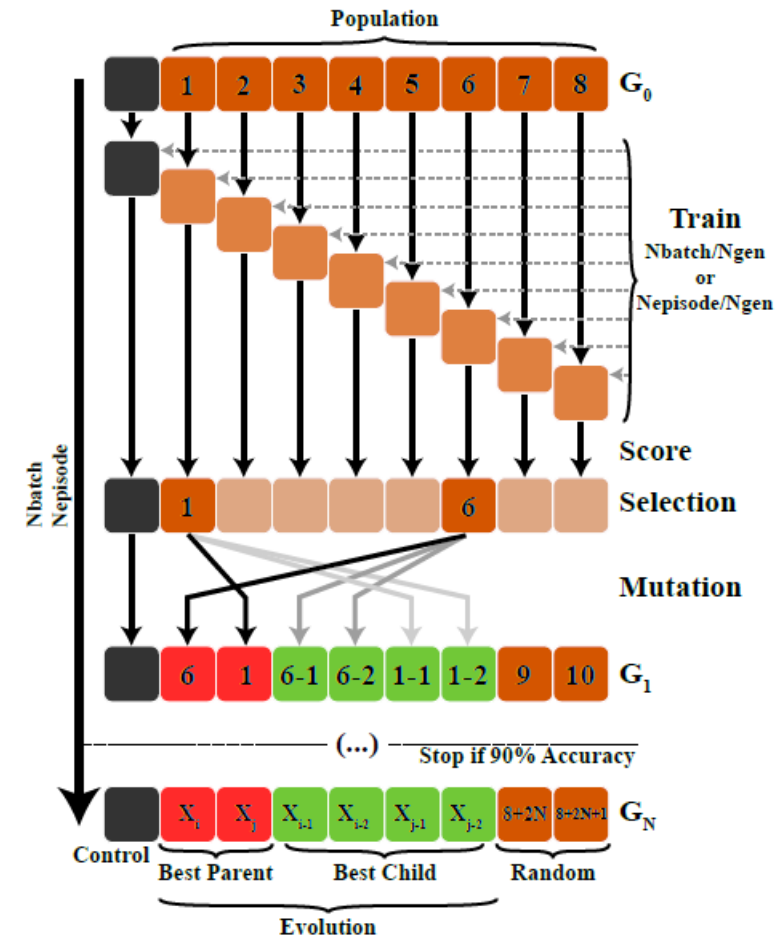
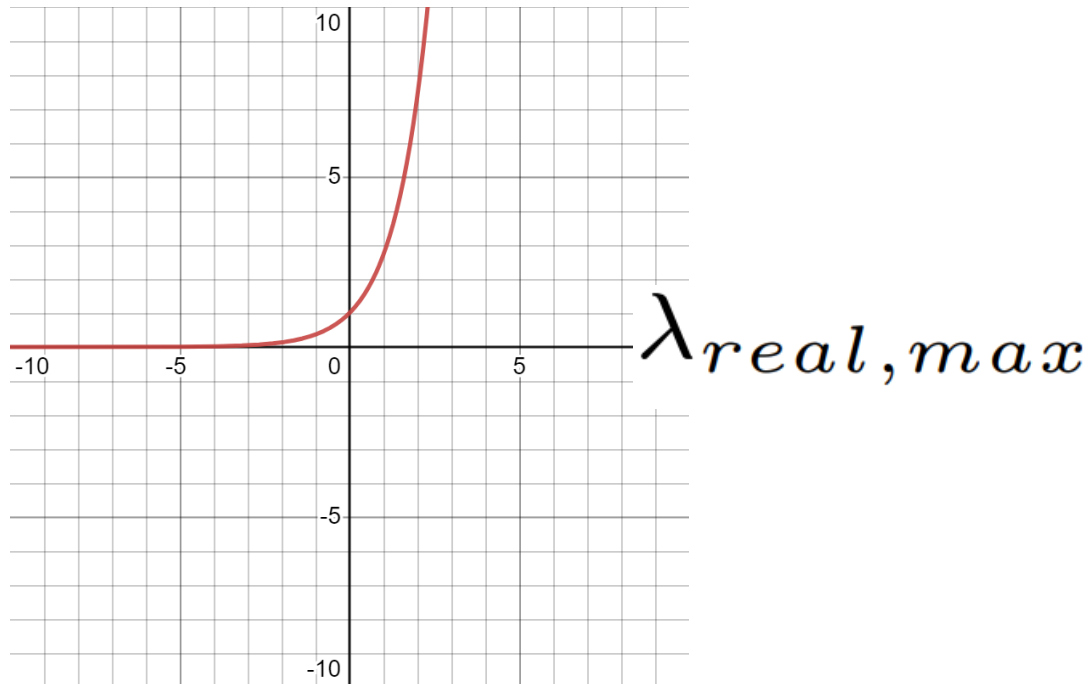
# Cost function for Stability informed learning

Stability term

$$CE + K(e^{\lambda_{real,max}})$$

Cross Entropy

Regularizer

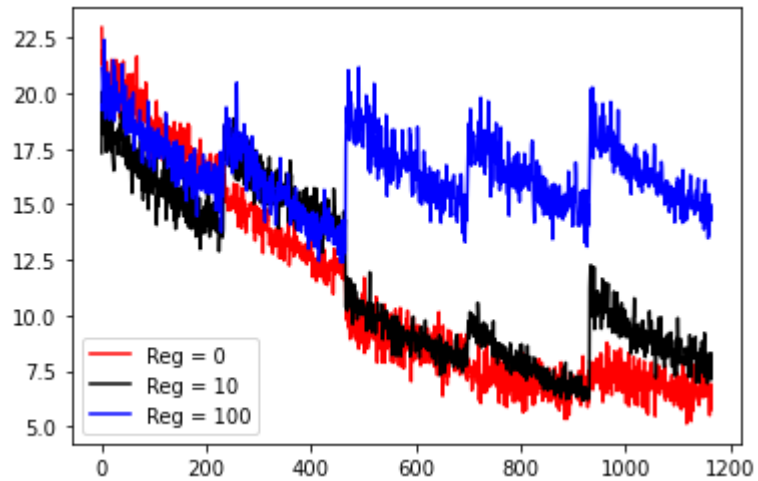


# *Structural convergence for stability informed evolution*

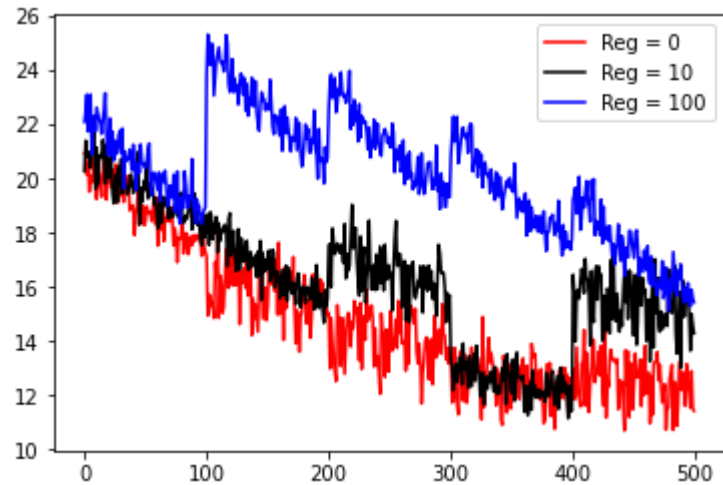
## CLASSIFICATION

## LOSS

**BATCHE SIZE = 30**



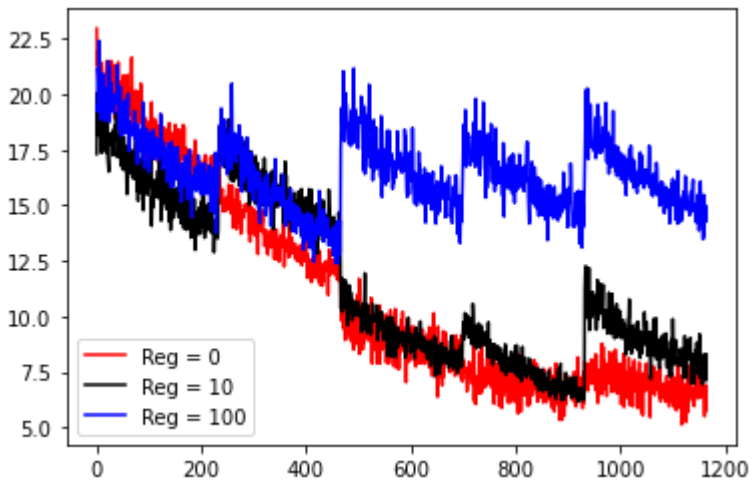
**BATCHE SIZE = 40**



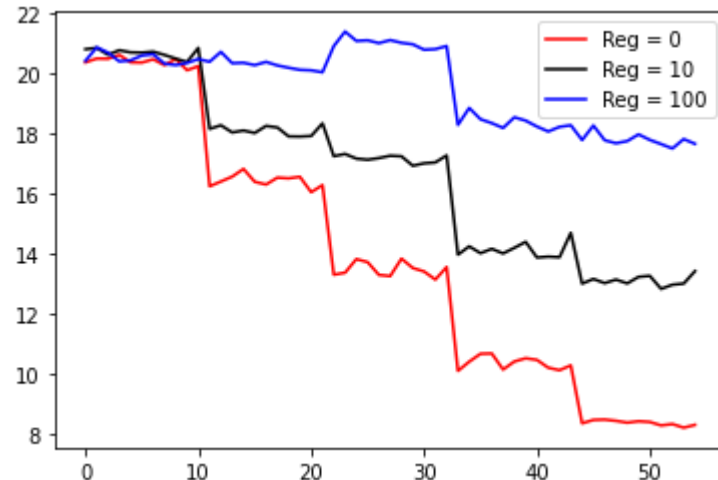
# Structural convergence for stability informed evolution

## CLASSIFICATION LOSS

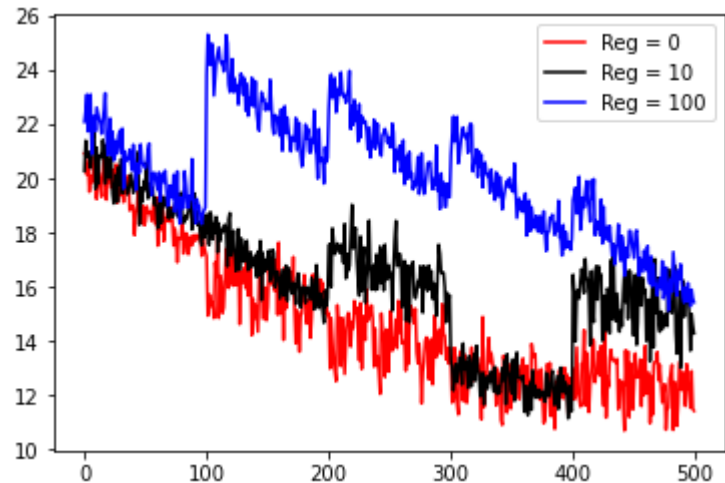
BATCHE SIZE = 30



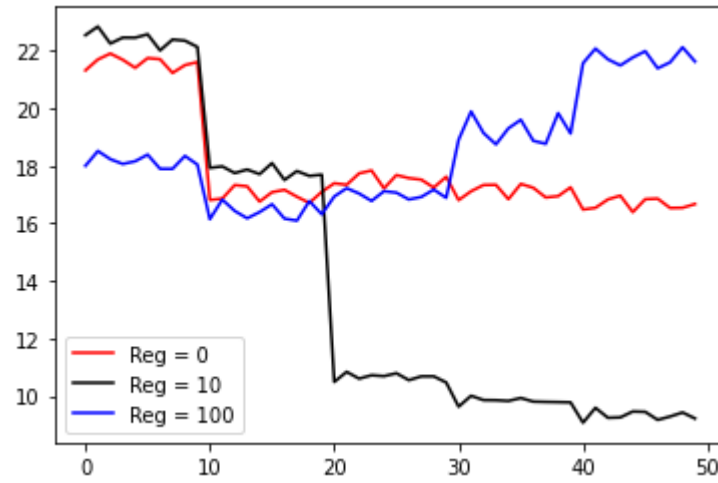
BATCHE SIZE = 500



BATCHE SIZE = 40



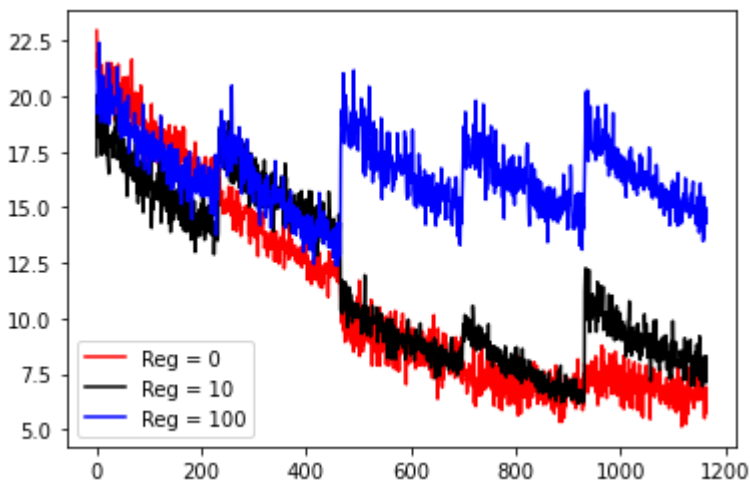
BATCHE SIZE = 700



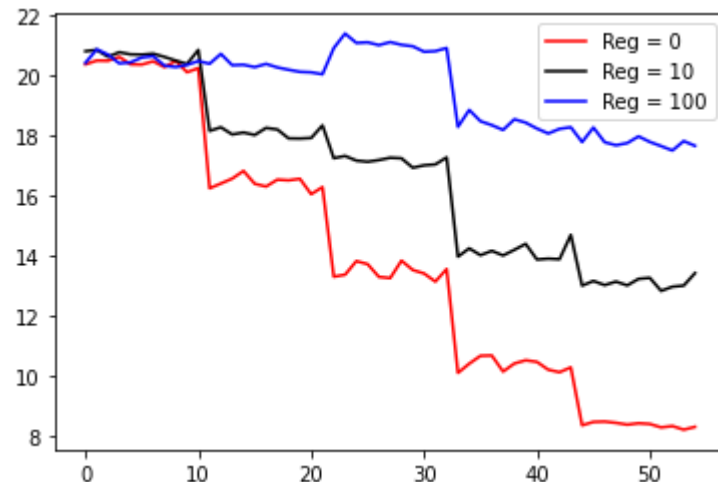
# Stability informed evolution outperformed SGD for a narrow batch size

## CLASSIFICATION LOSS

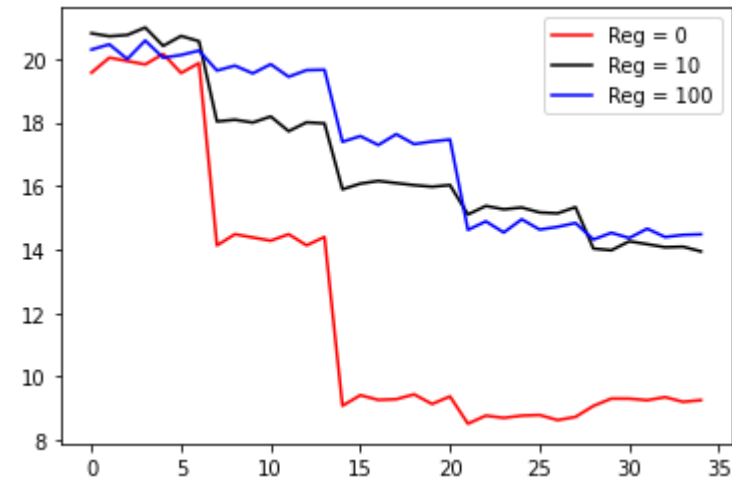
BATCH SIZE = 30



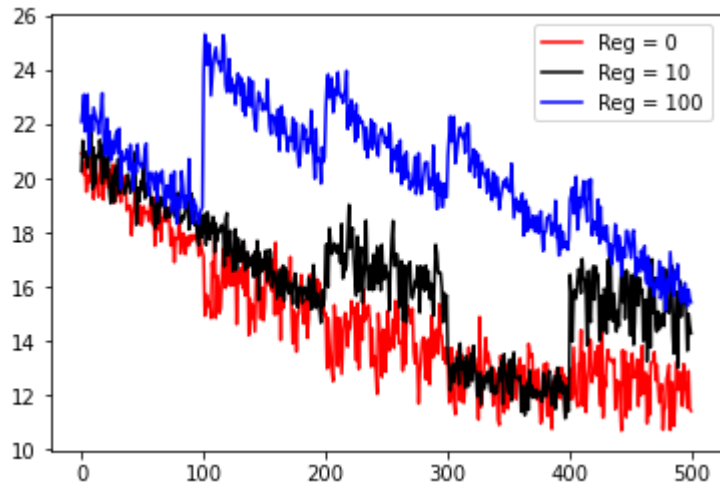
BATCH SIZE = 500



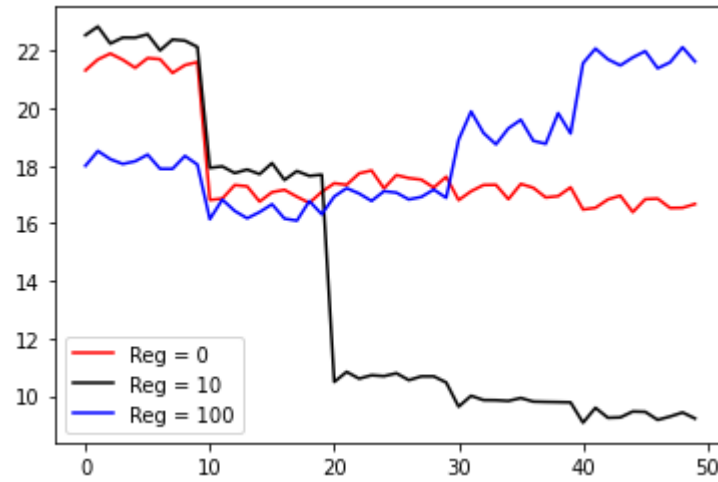
BATCH SIZE = 1000



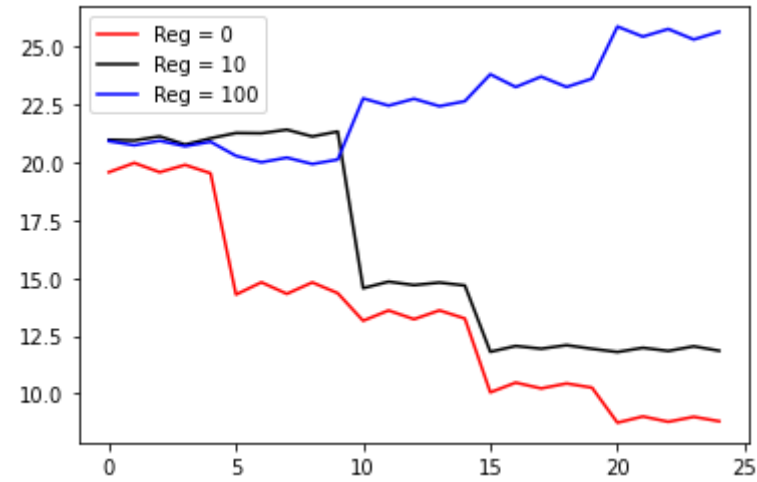
BATCH SIZE = 40



BATCH SIZE = 700



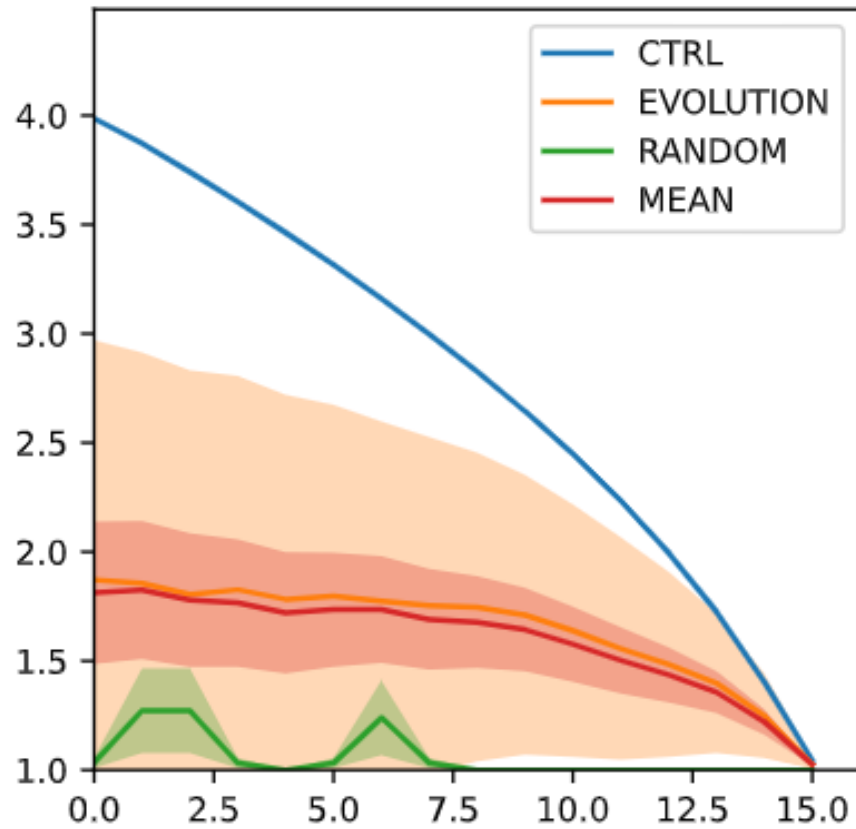
BATCH SIZE = 1300



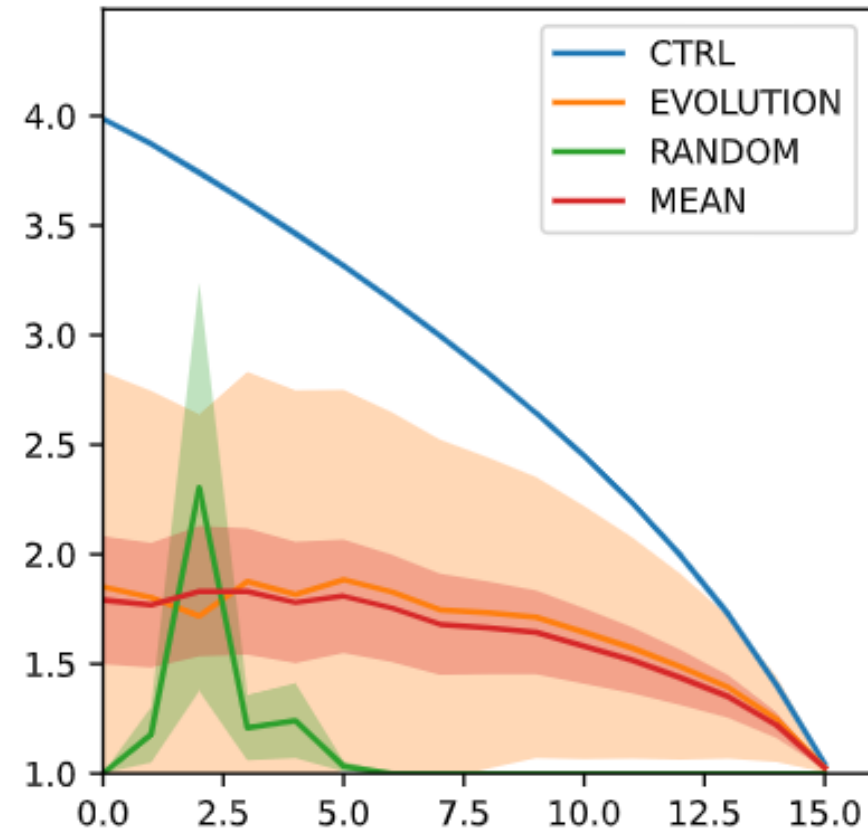
BATCHES

# *Evolutionary Stability is not enhanced by Network Stability*

Reg = 0  
MNIST



Reg = 1  
MNIST



# *Conclusions*

- (i) Stability informed evolutionary learning outperformed standard SGD for a narrow batch-size
- (i) Stability informed evolutionary learning did not produce more robust lineages

Thank you!