Artificial Neural Networks Functionalized By Evolution

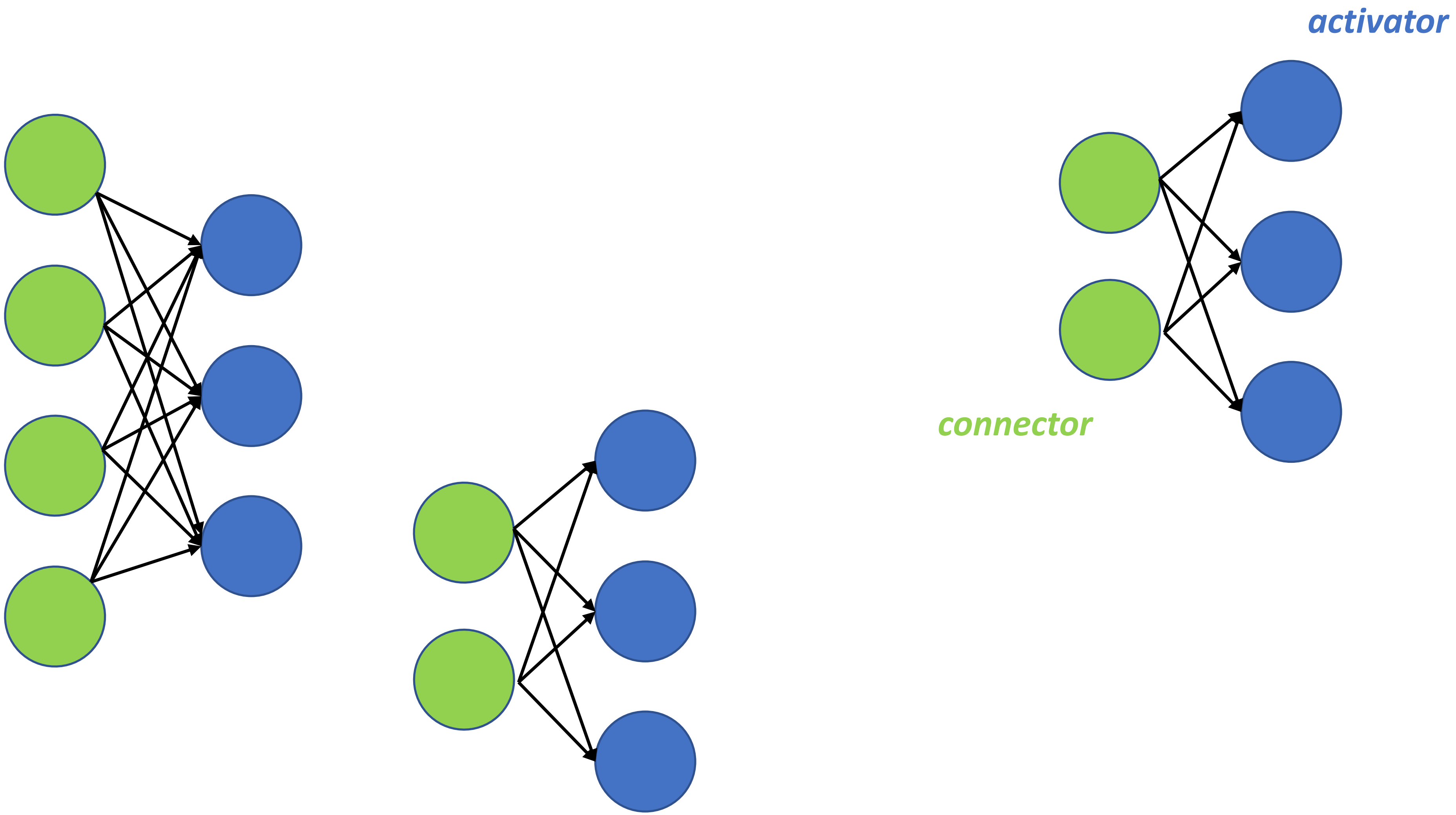
Furfaro, Fabien, Avner Bar-Hen, and Geoffroy Berthelot. "An Artificial Neural Network Functionalized by Evolution." *arXiv preprint arXiv:2205.10118* (2022).

Presented By
Dennis Manjaly Joshy, ME
Yi Lab, UCSB

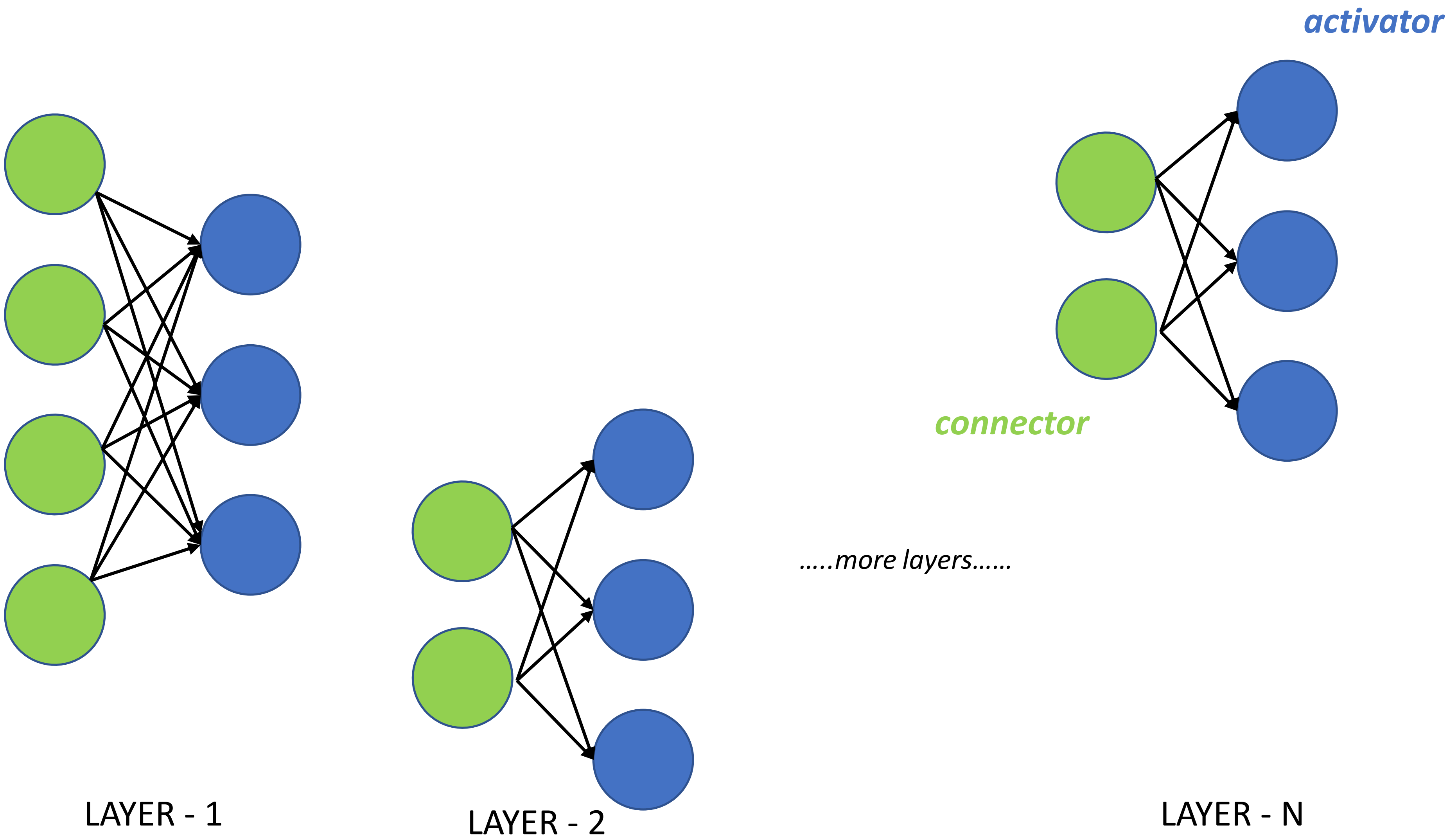
Outline

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

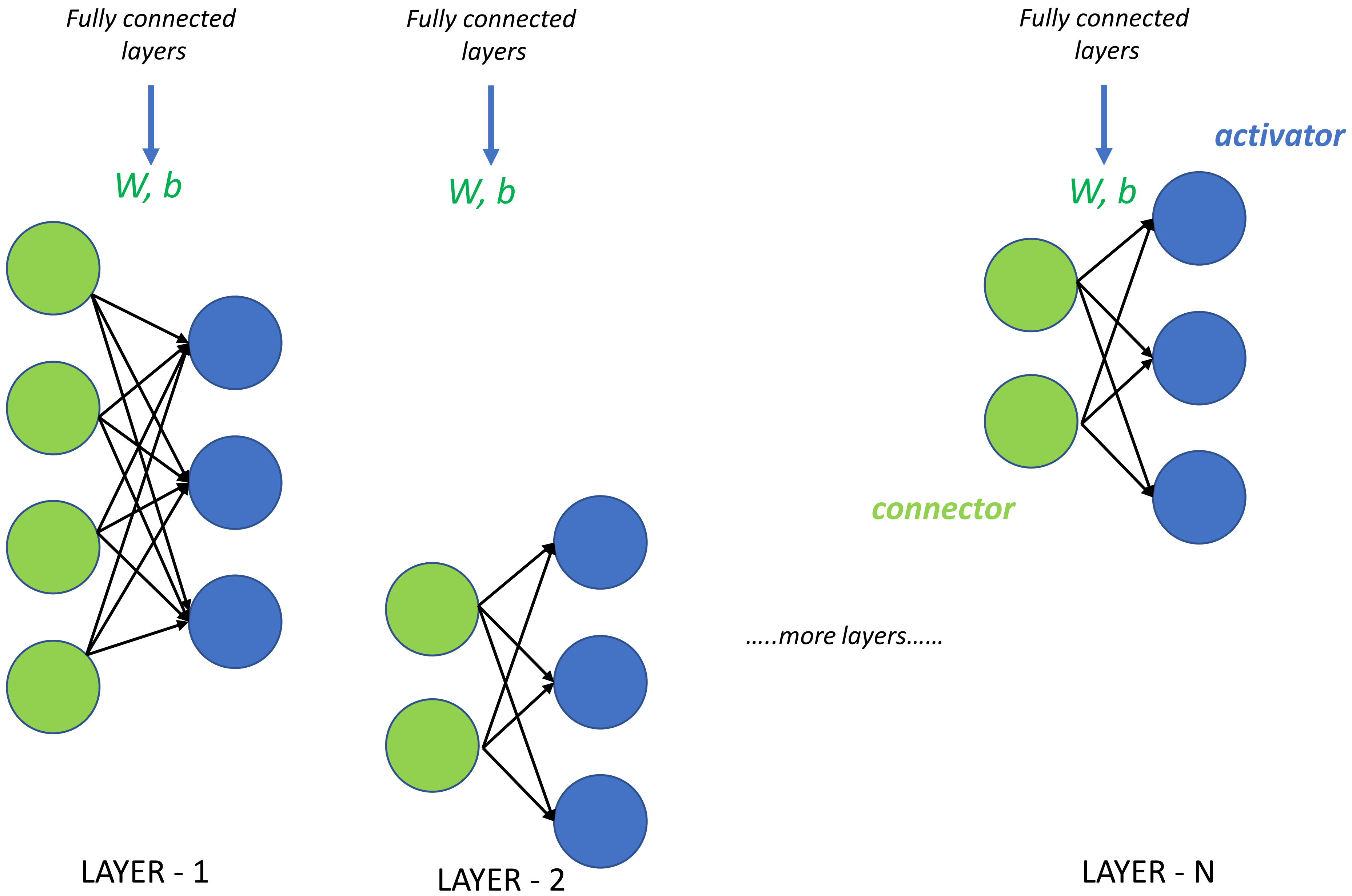
A more 'Neuro-Inspired' Topology



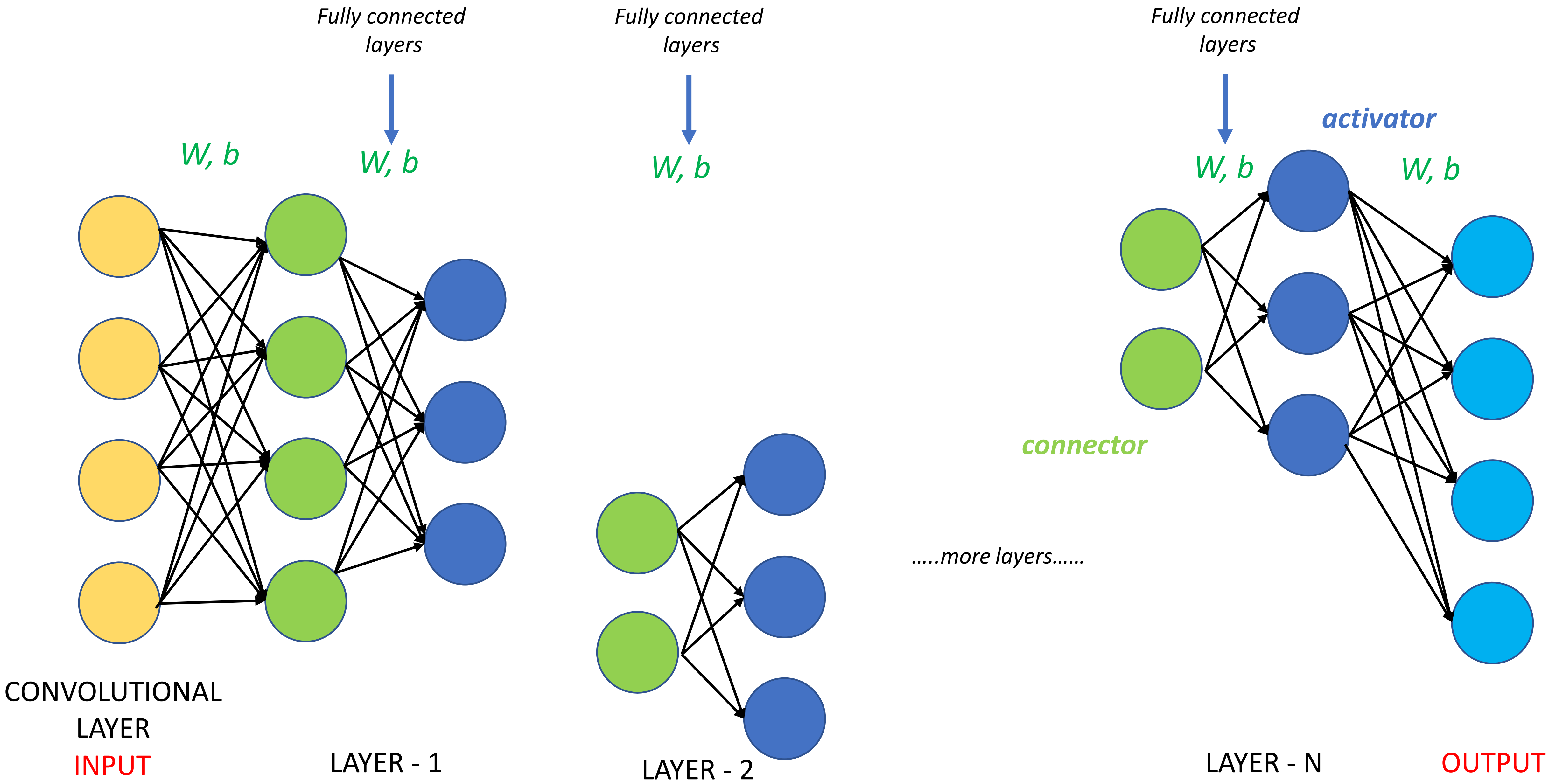
A more 'Neuro-Inspired' Topology



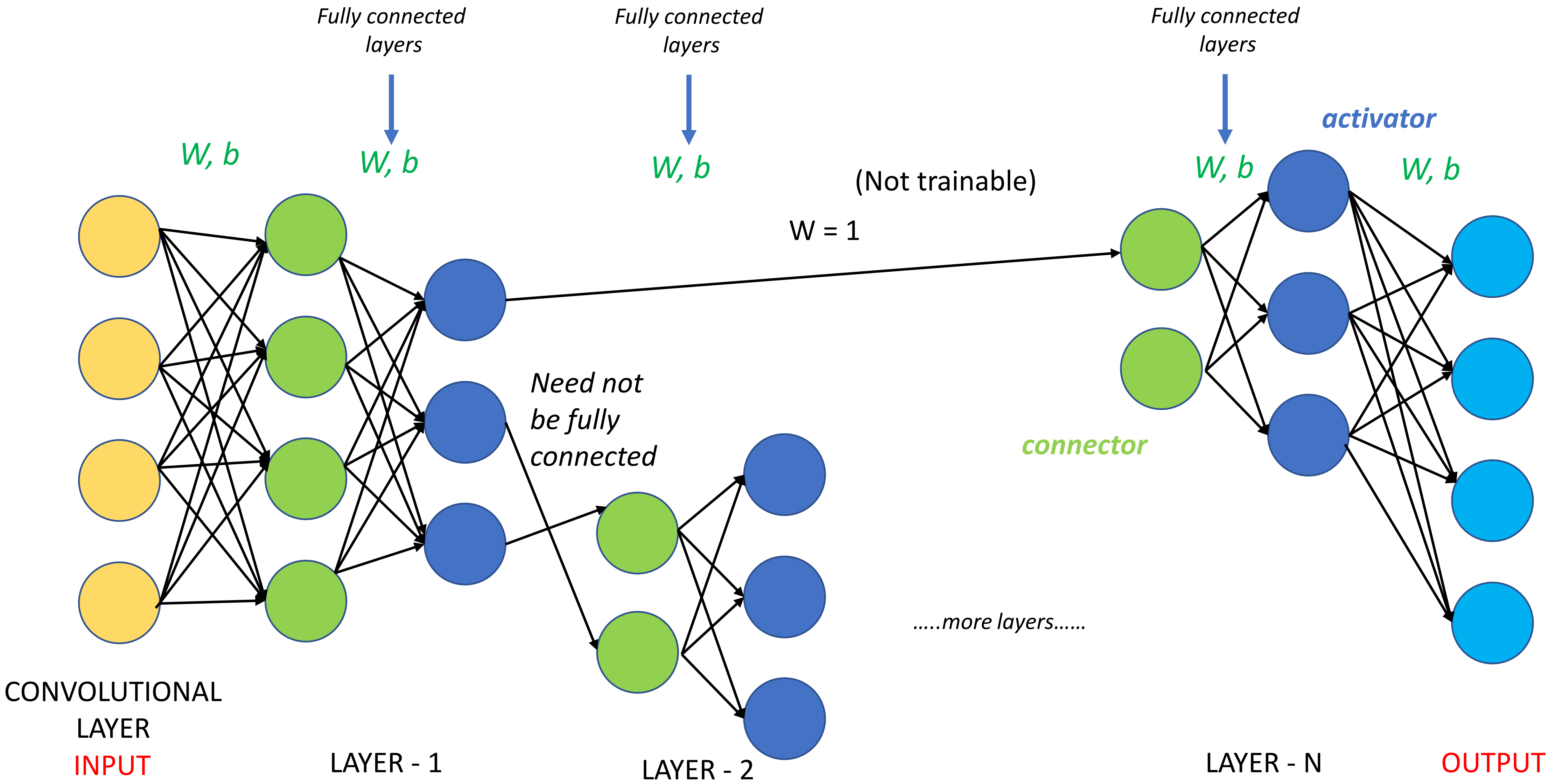
A more 'Neuro-Inspired' Topology



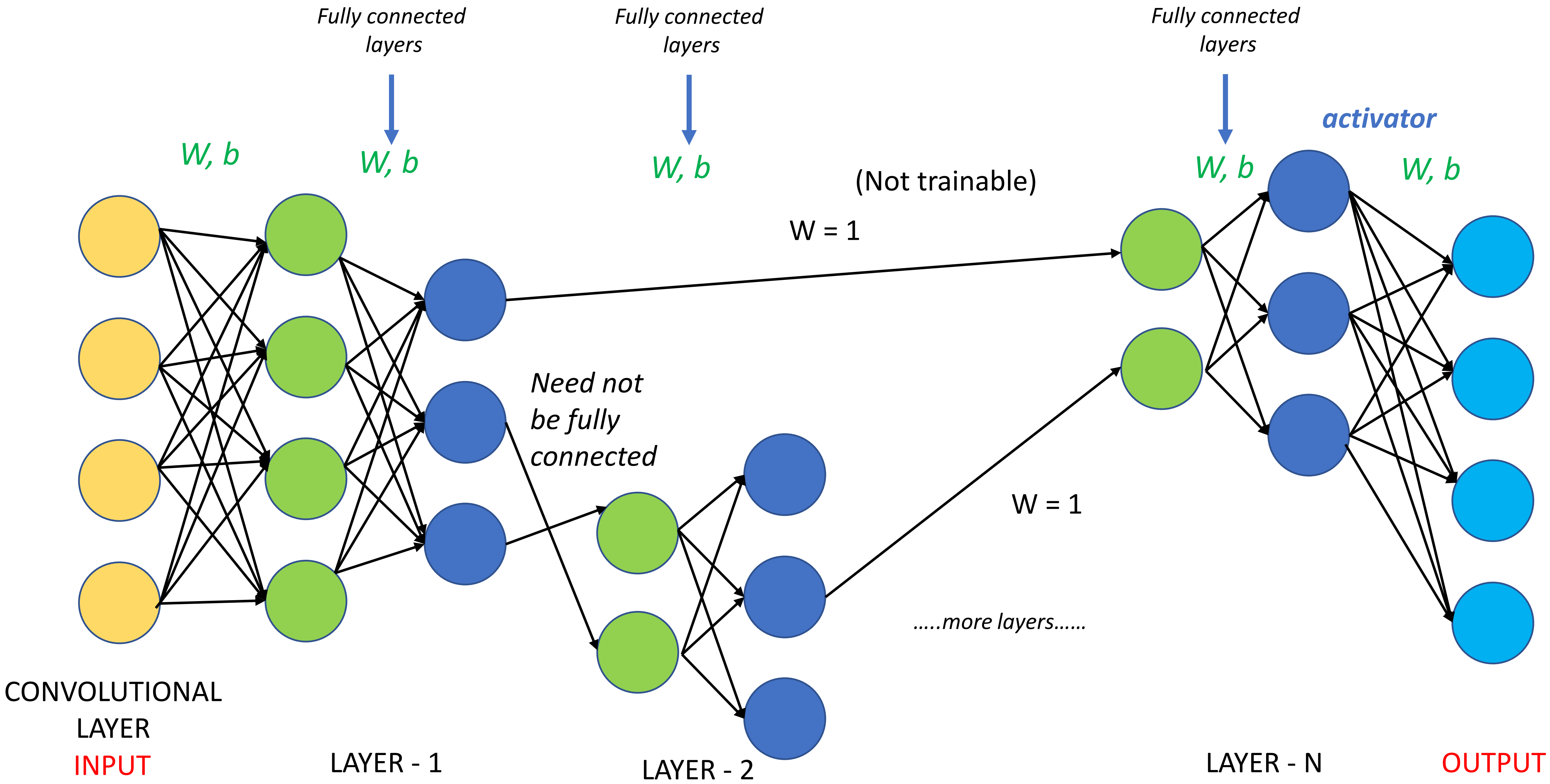
A more 'Neuro-Inspired' Topology



A more 'Neuro-Inspired' Topology

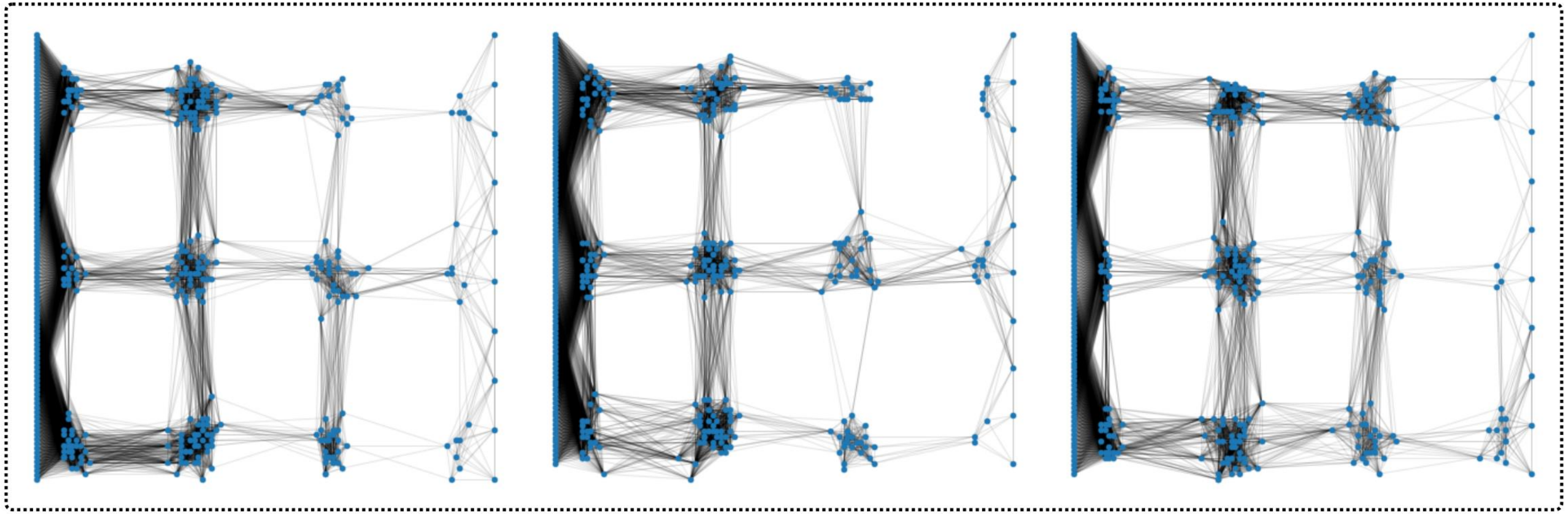


A more 'Neuro-Inspired' Topology



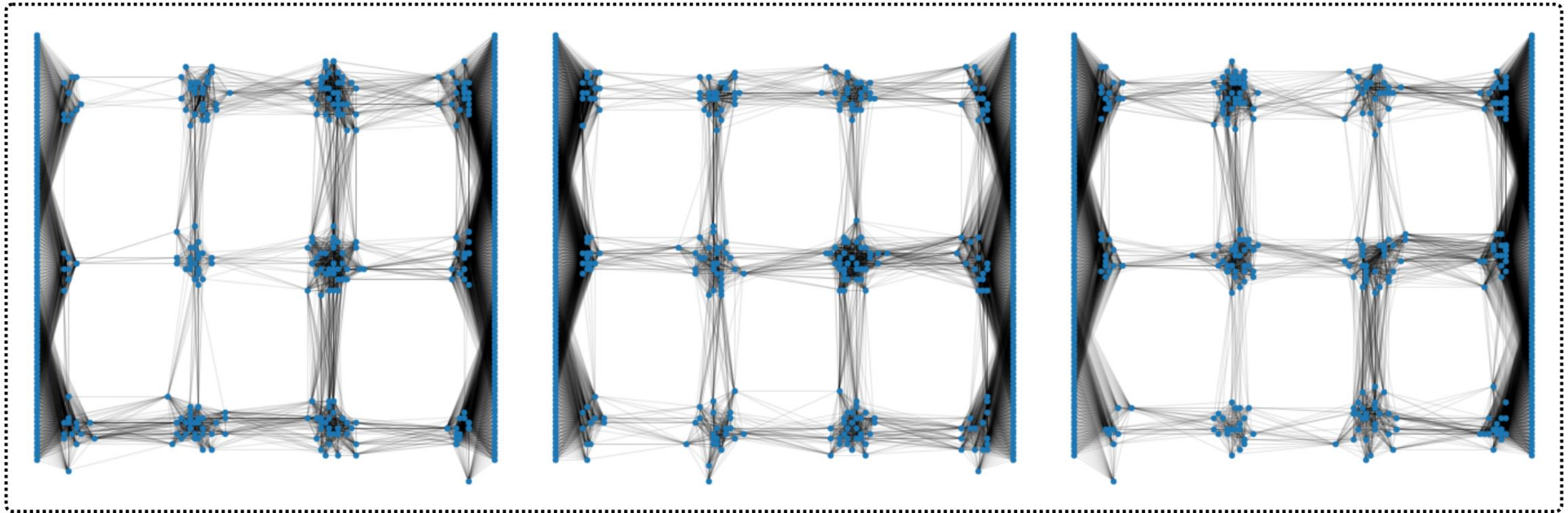
Visualizing the Networks Made with these rules

Taille In ≠ Out



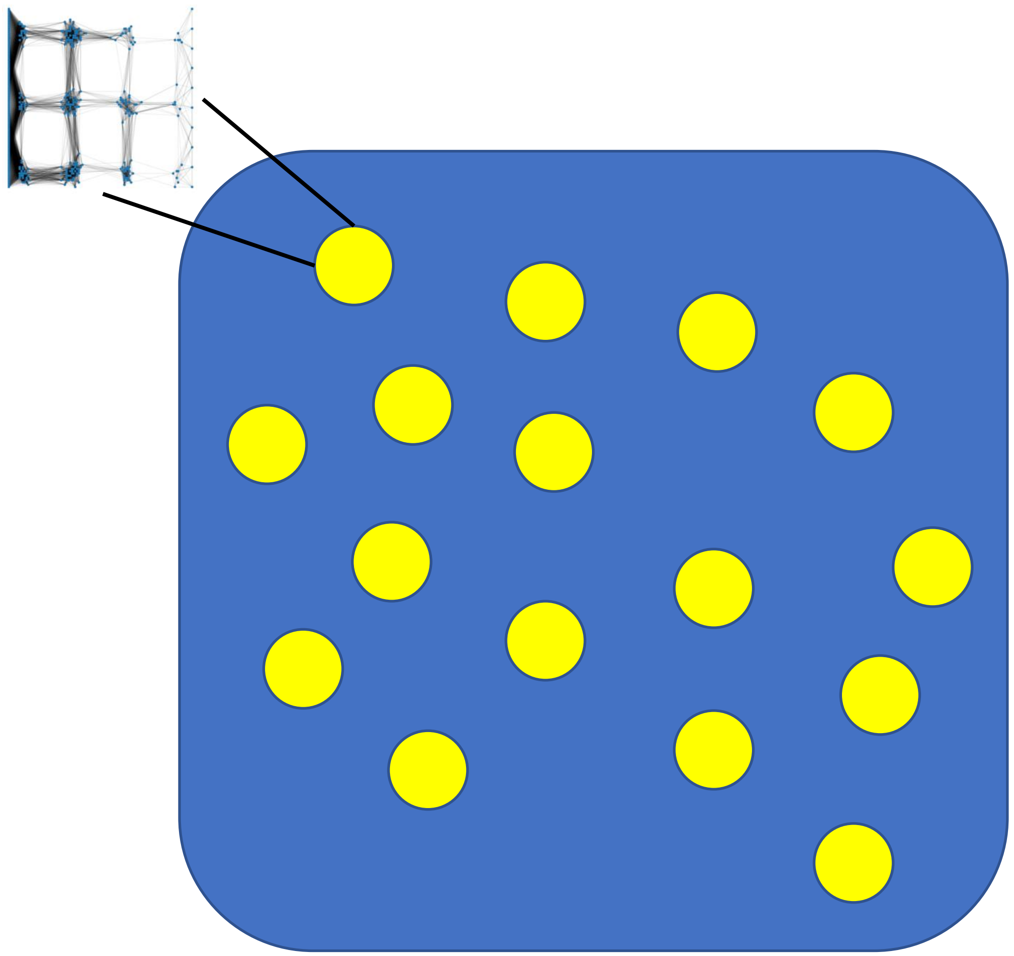
<u>Applications</u> :
Base MNIST / IMdb
Prise de décision

Taille In = Out



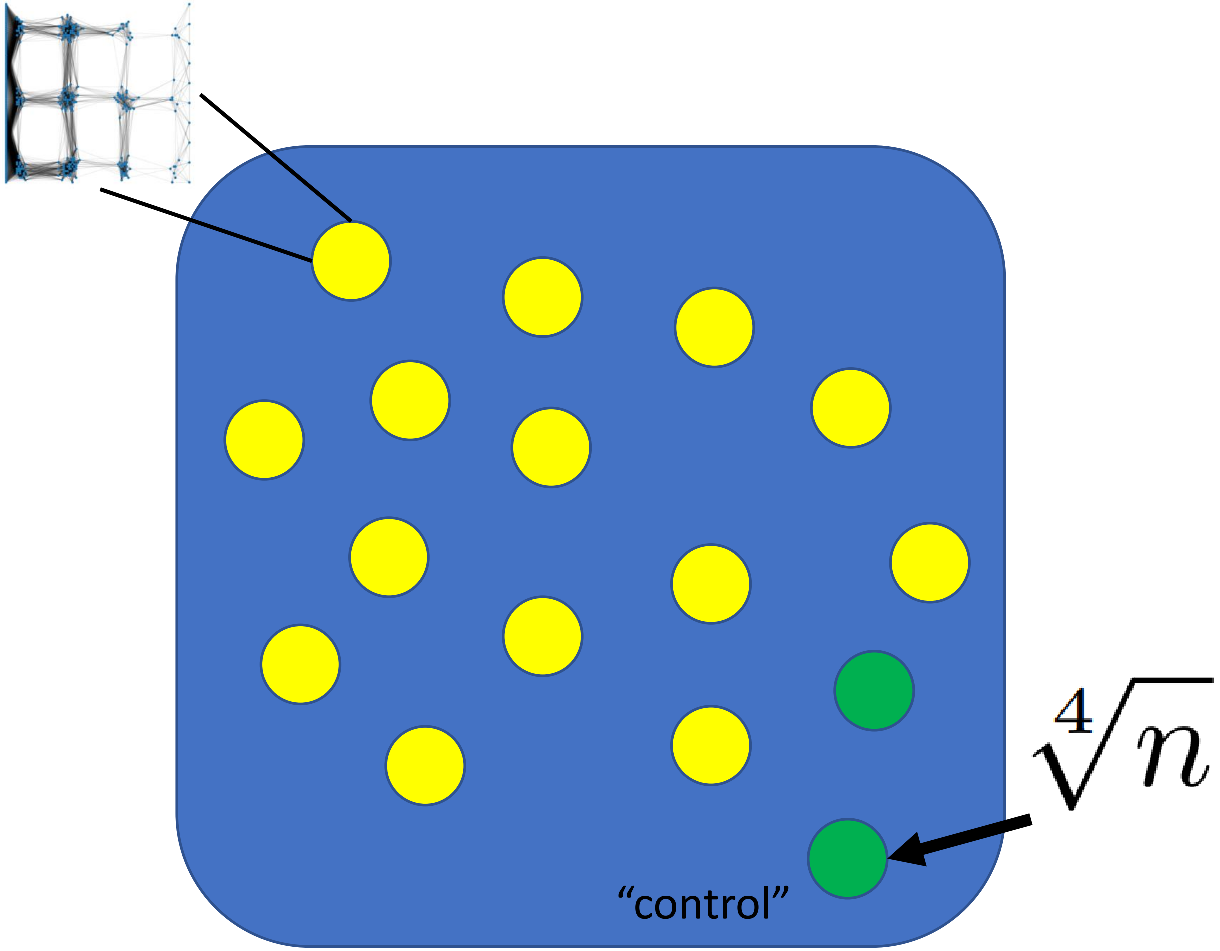
<u>Applications</u> :
Base ImageNet / Pascal2
Prédiction
Detection d'objet

Functionalizing the Network using Evolution (Pseudo-Darwinism)



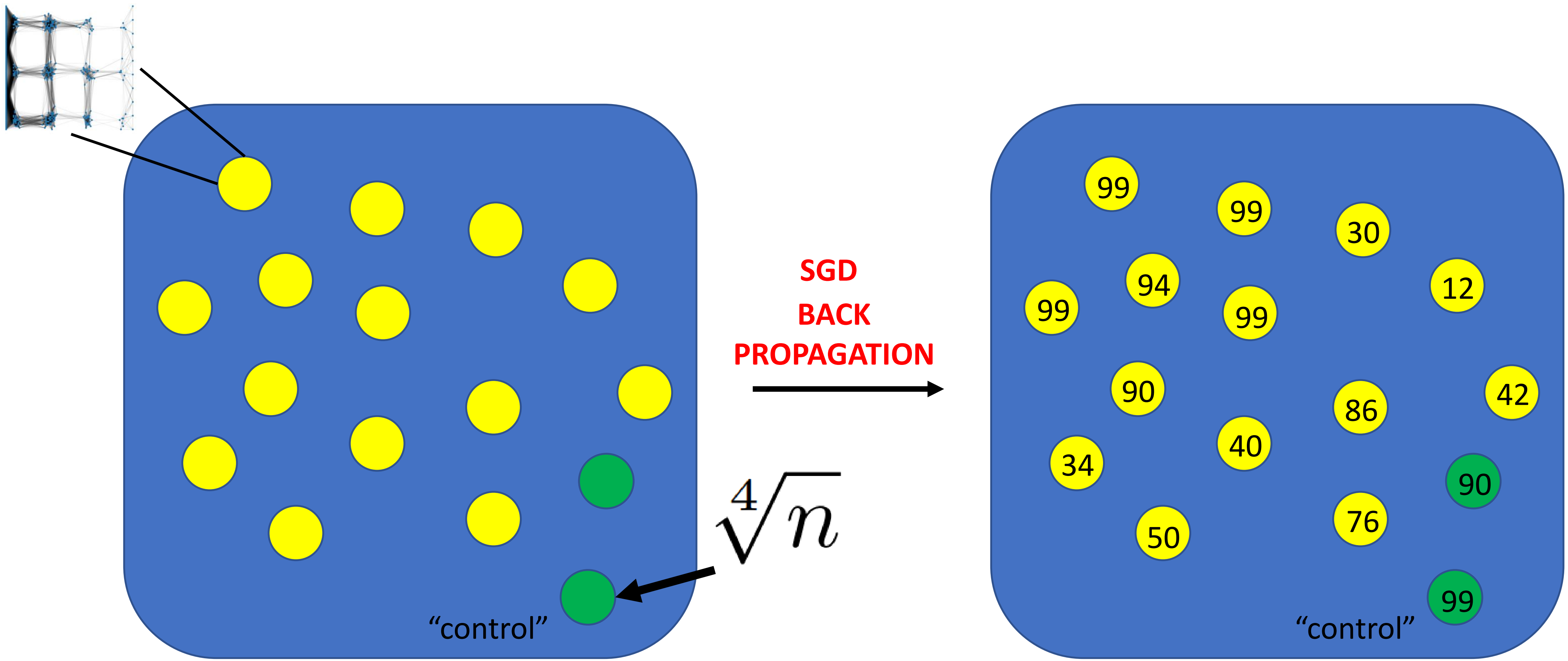
$$n = 16$$

Functionalizing the Network using Evolution (Pseudo-Darwinism)



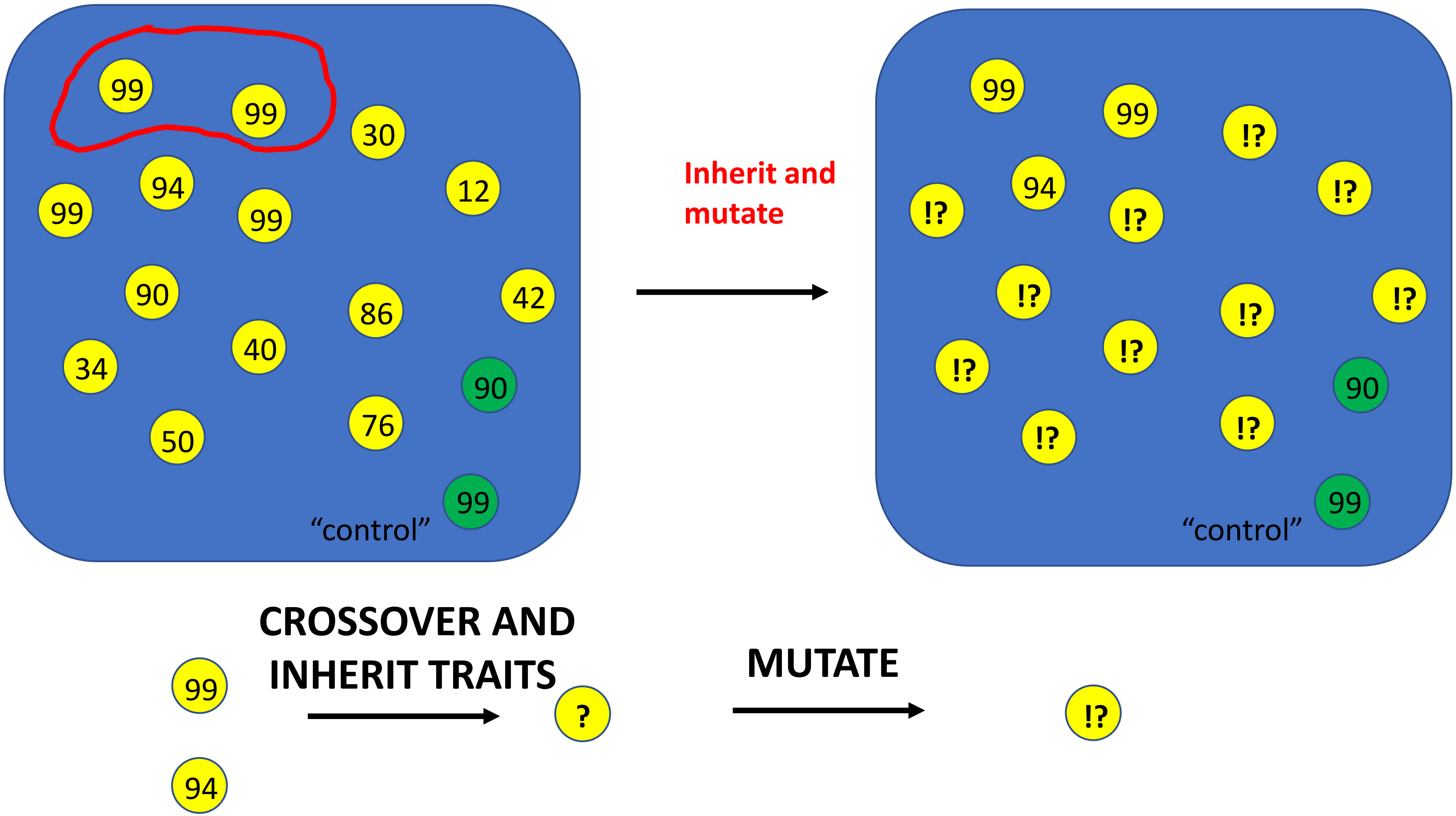
$$n = 16$$

Functionalizing the Network using Evolution (Pseudo-Darwinism)

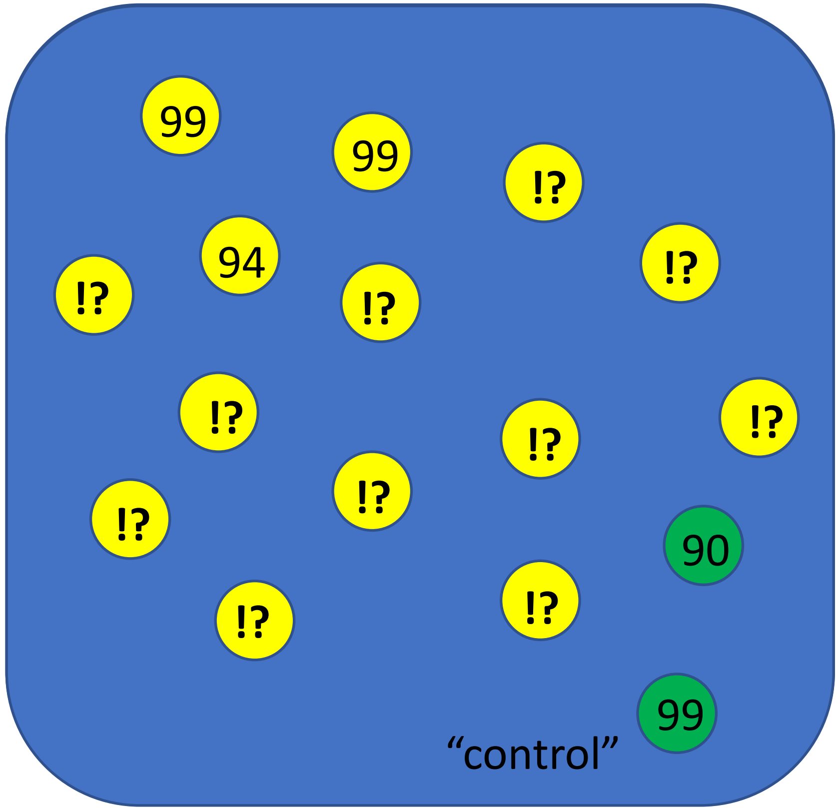


$$n = 16$$

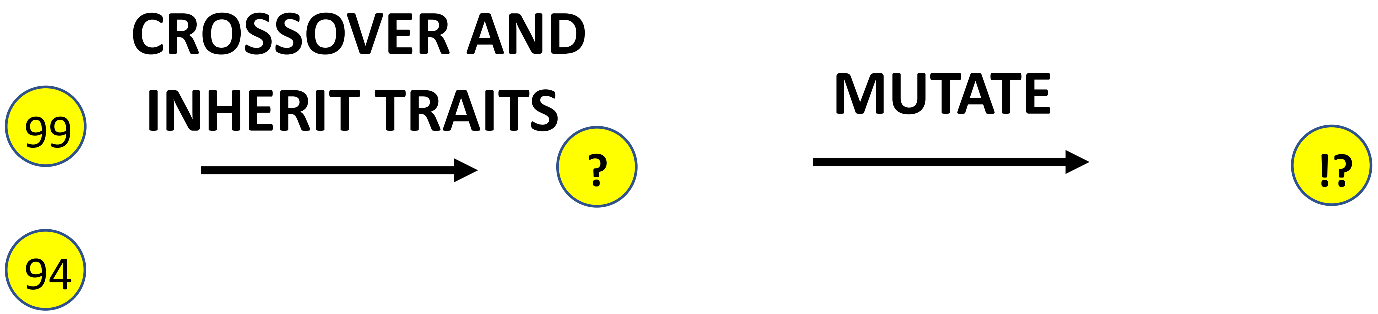
Functionalizing the Network using Evolution (Pseudo-Darwinism)



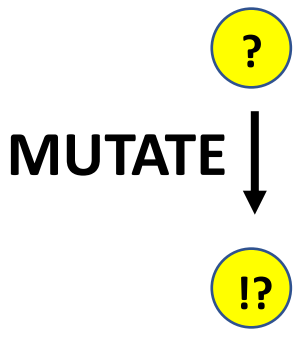
Functionalizing the Network using Evolution (Pseudo-Darwinism)

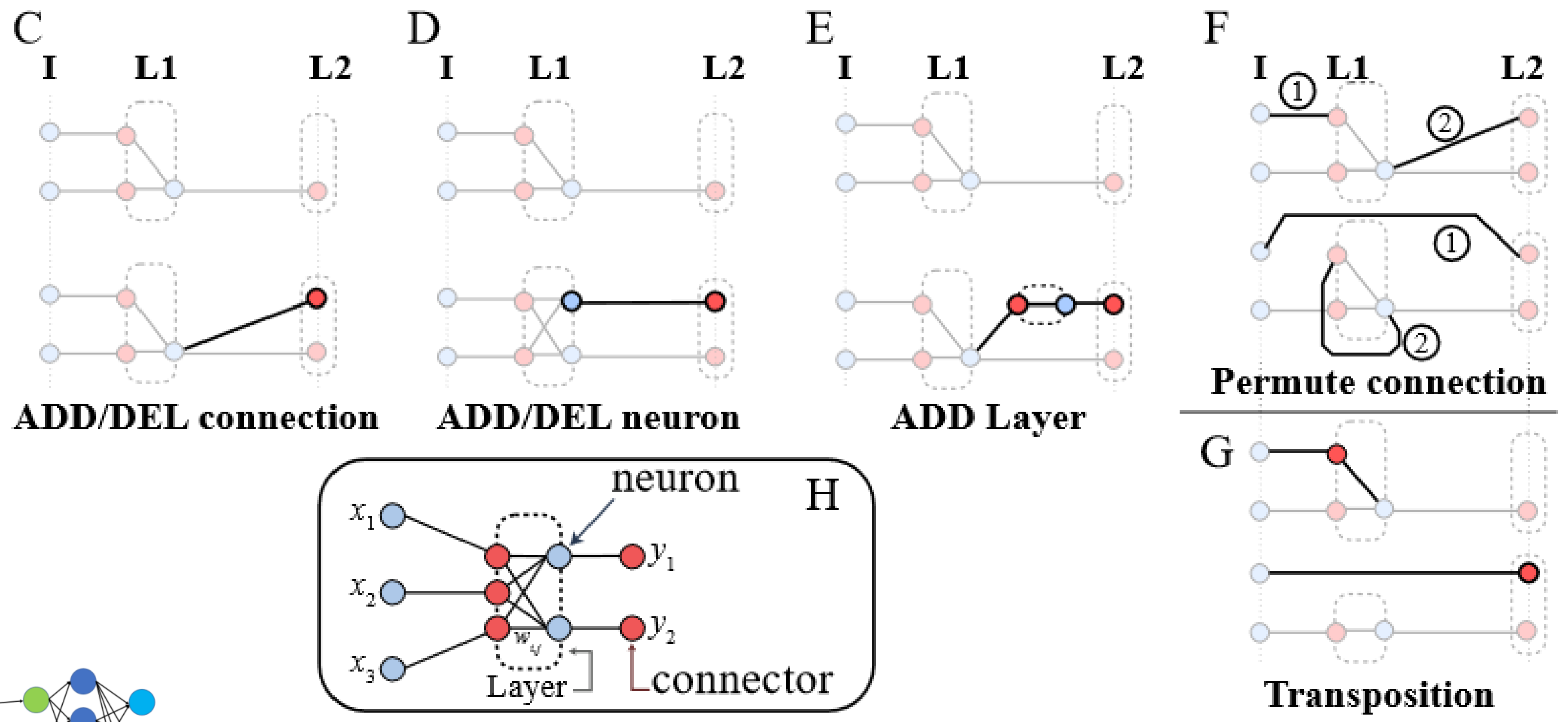


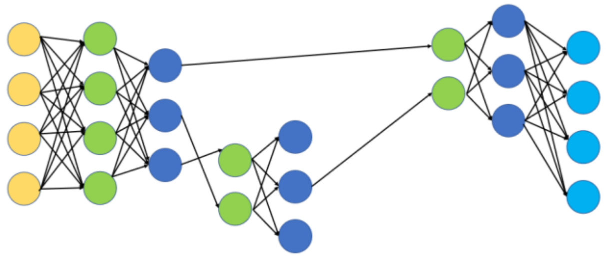
3	**Parents from G - 1**	$\alpha = \sqrt{n} - 1$
9	**Offspring (inherited + mutated)**	α^2
2	**New Networks (random)**	α
2	**Control**	$\sqrt[4]{n}$



All Possible Mutations to change network topology



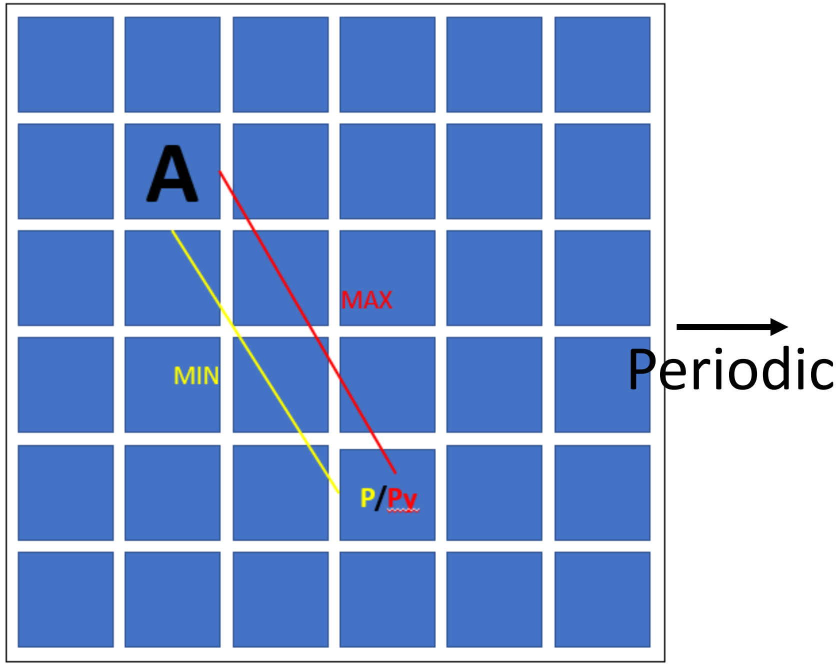




Algorithm summary

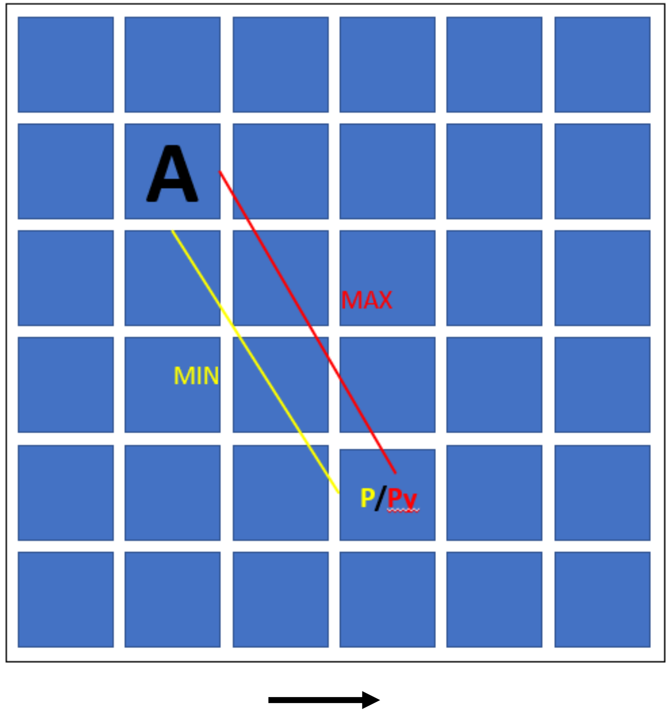
1. draw n random realizations of neural networks, from which $\lfloor \sqrt[4]{n} \rfloor$ will serve as control networks. This correspond to the initial population networks at the first generation (G_0)
2. train the current population using gradient descent and a number of batches
3. select the α most 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 α^2 independent children, inheriting the structure (traits) from the selected parents (plasticity steps) networks and mutate their structures
5. insert α 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



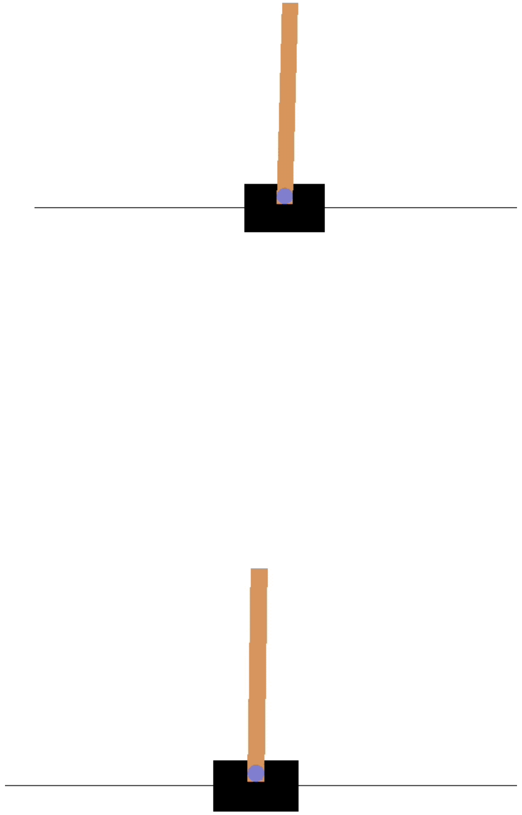
Tasks used to benchmark learning algorithm

TAG

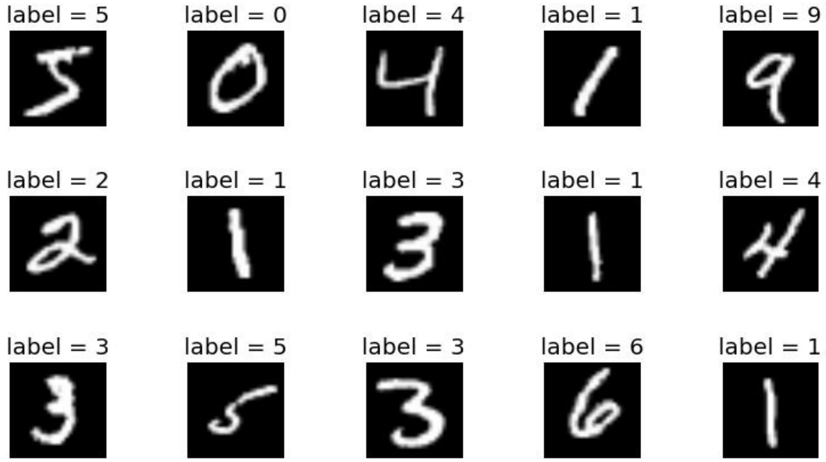


Periodic

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

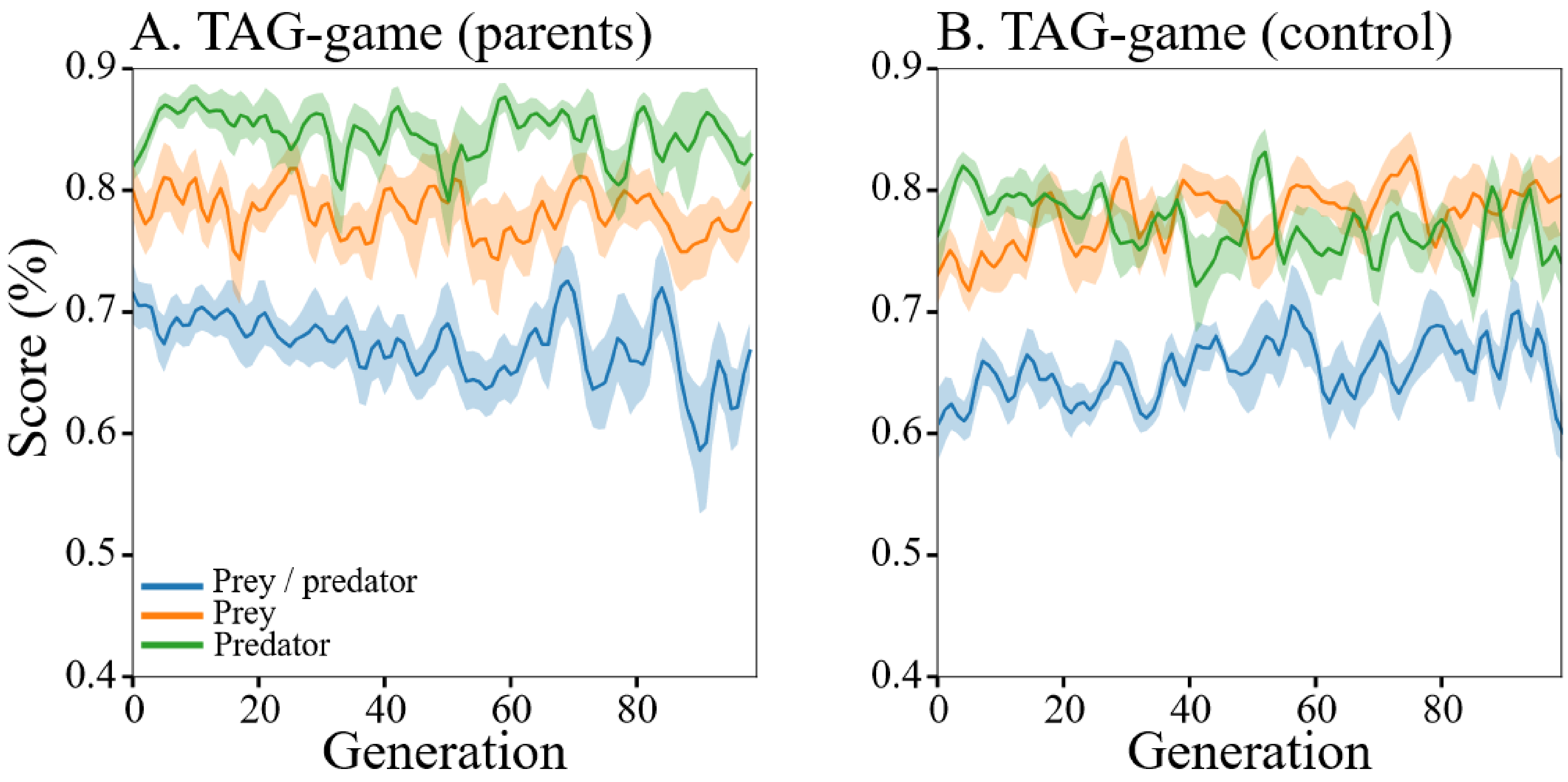
12800 episodes

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

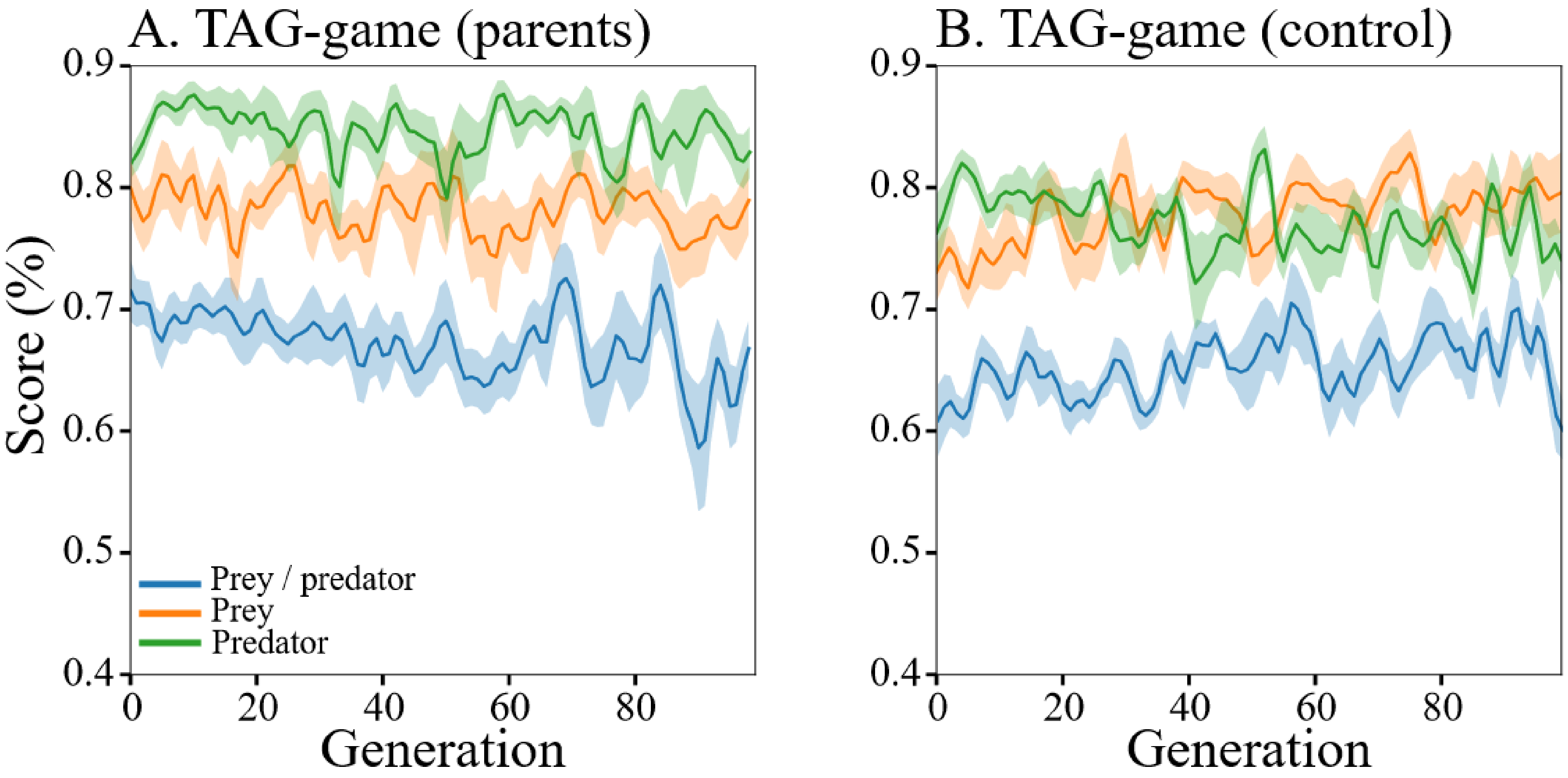
12800 episodes

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

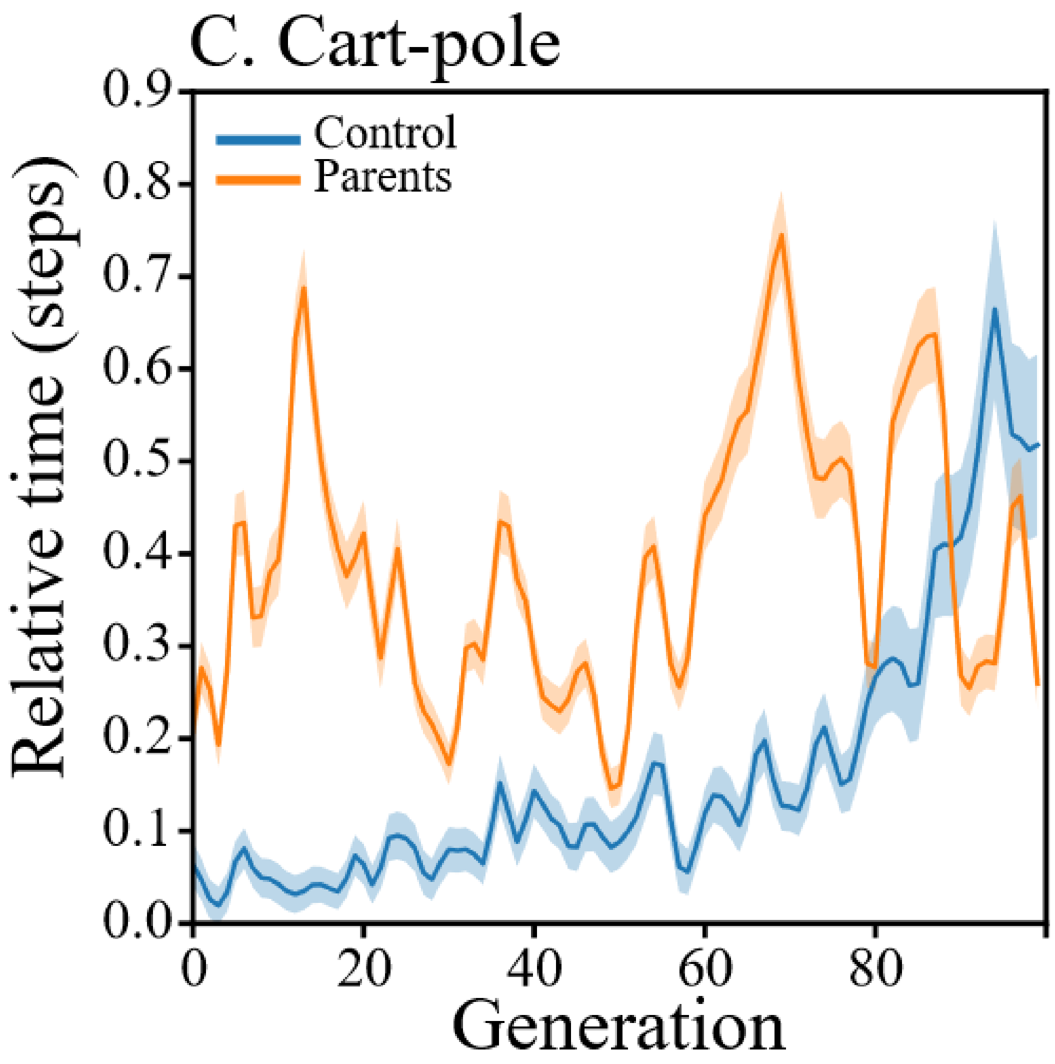
25000 episodes

100 generations

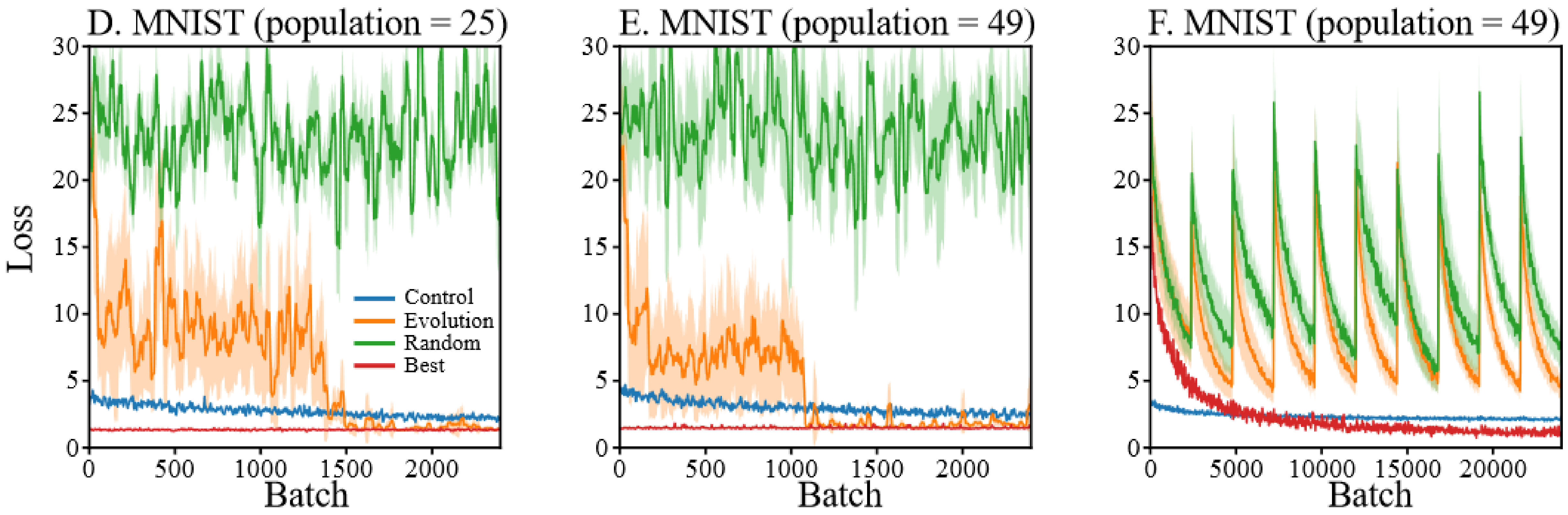
250 episodes/gen

300 time steps/episode

1 batch (for SGD) is 25 time steps

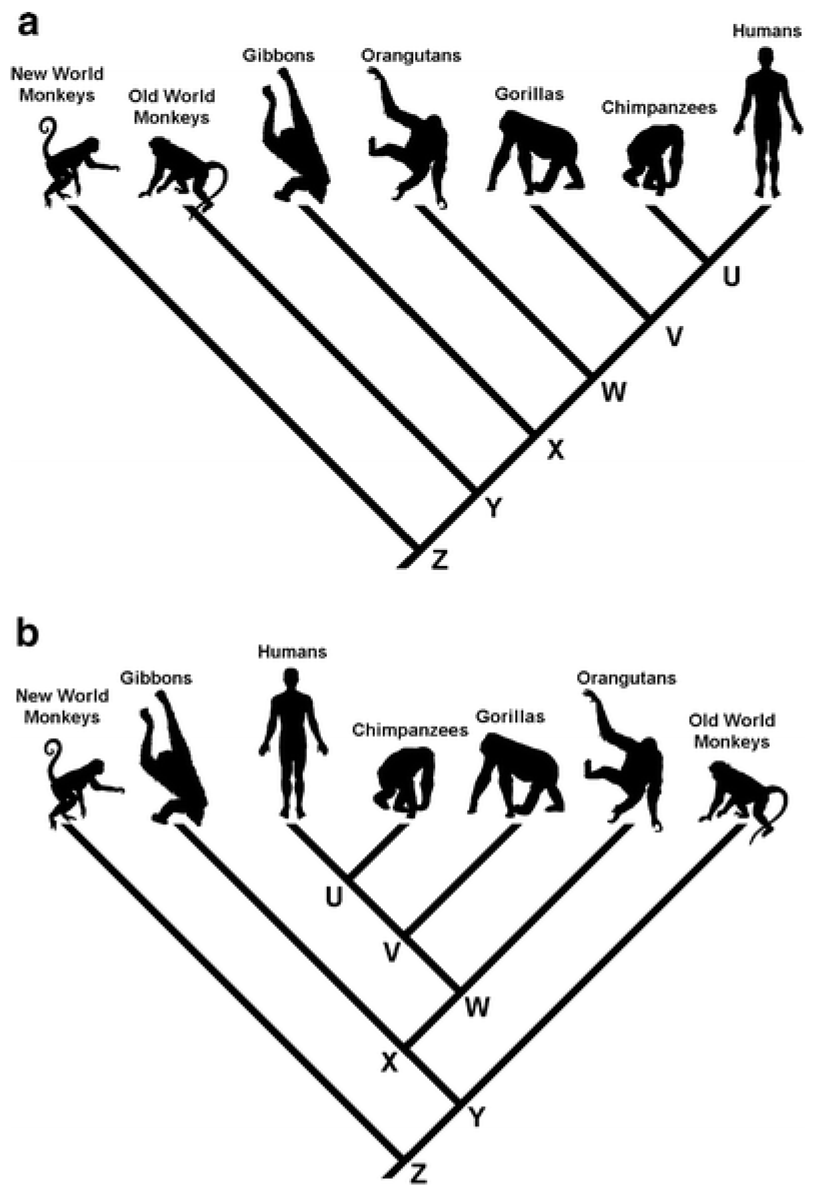


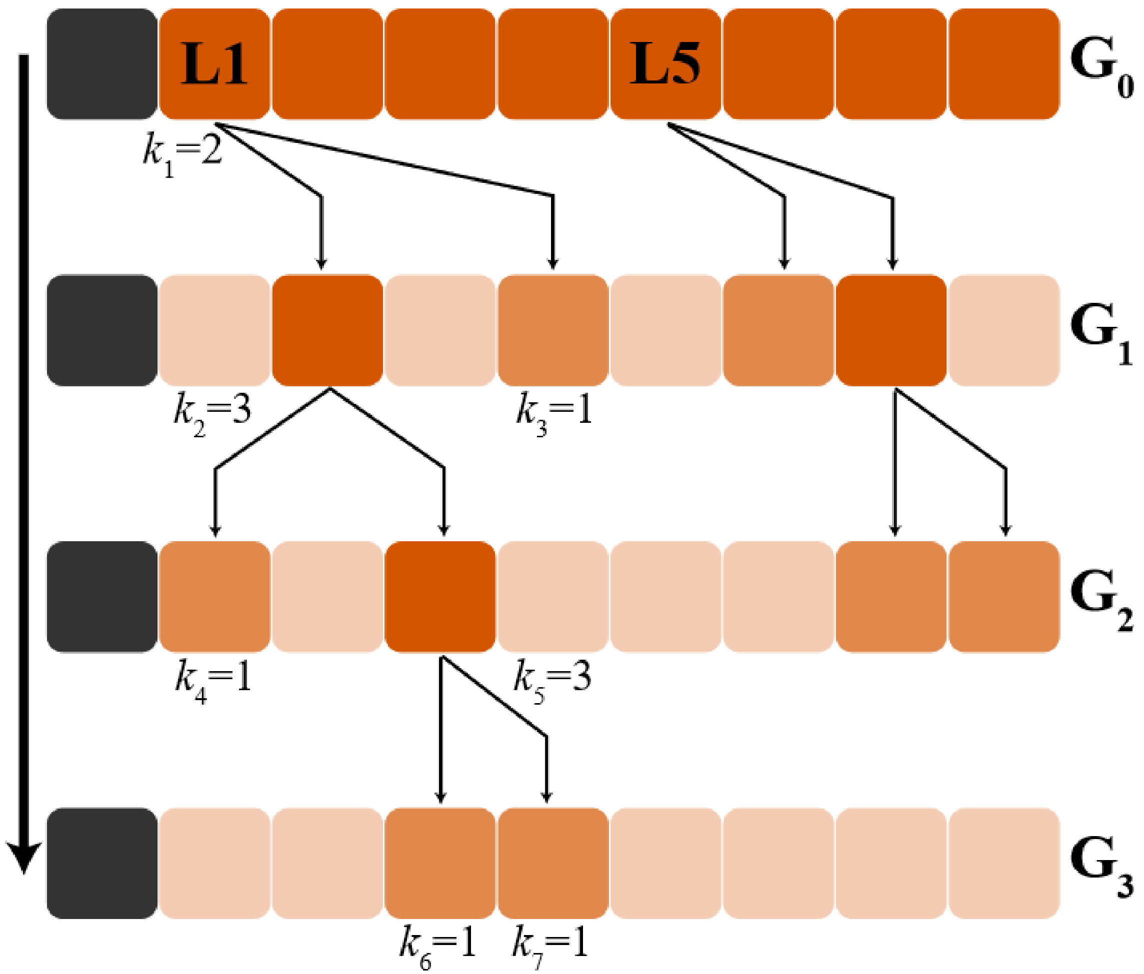
MNIST Classification loss vs batch-size



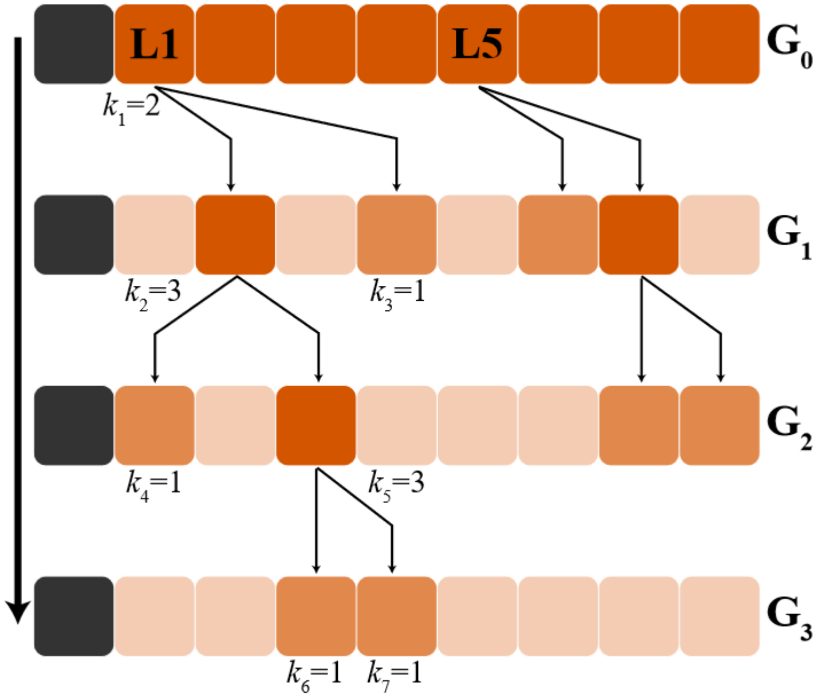
When this 'structural convergence' occurs, the evolutionary networks become more efficient than both the control and the random ones. The structural convergence occurs faster when we have a population of 49 networks (1100 batches) than when we have a population of 25 networks (1500 batches).

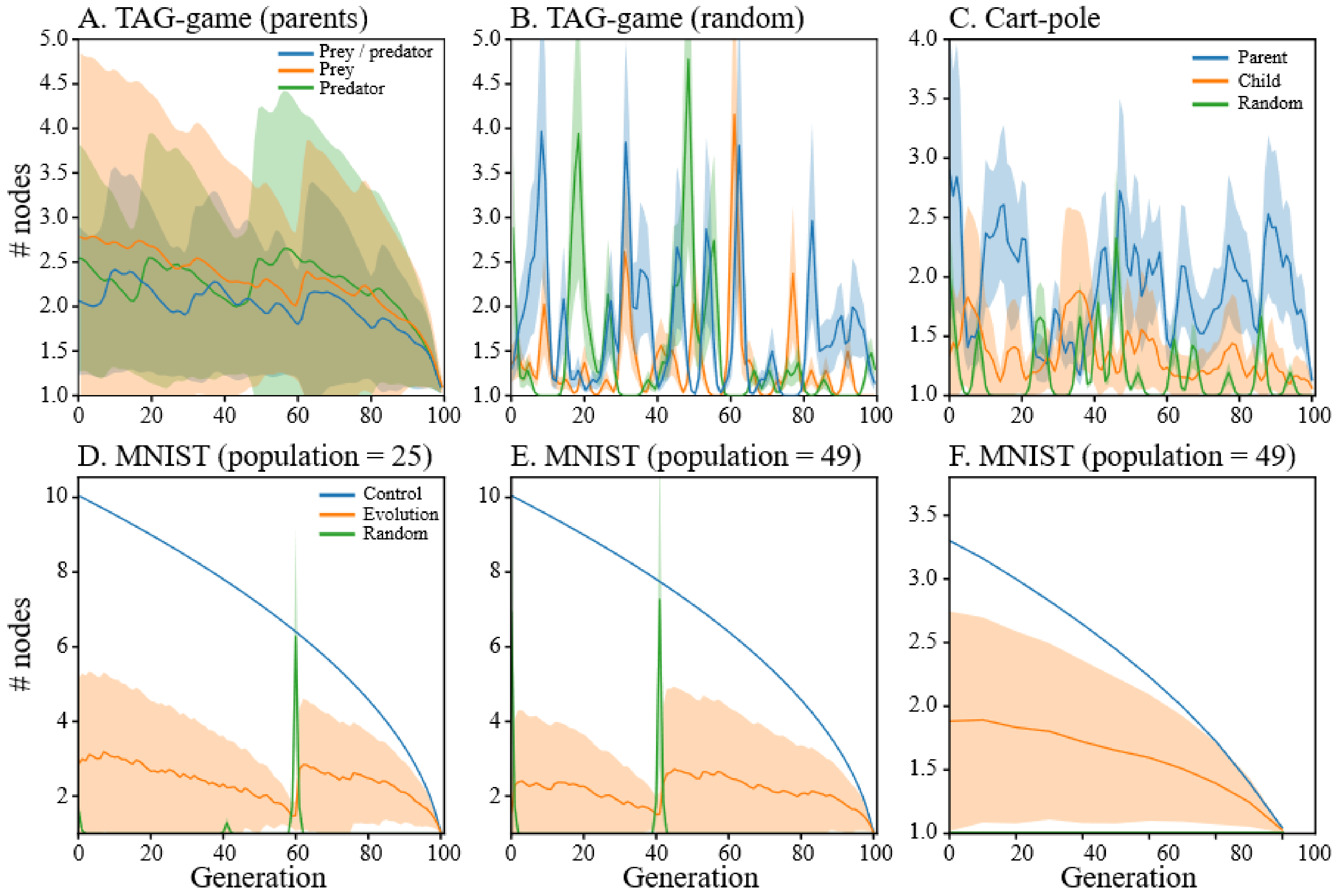
Evolutionary Stability of a lineage in the population





Evolutionary Stability of a lineage in the population





Conclusions

(i) the larger the network, the more evolutionarily stable it is.

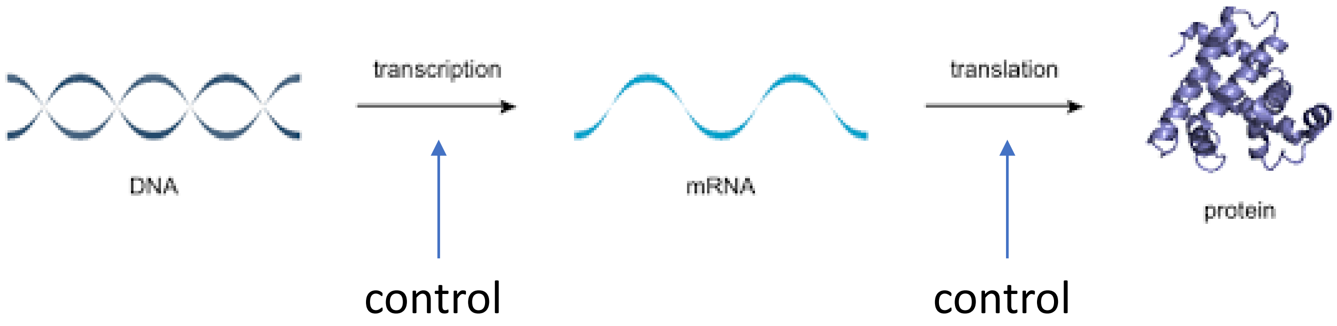
(i) the greater the number of individuals in the initial population, the faster it converges.

(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

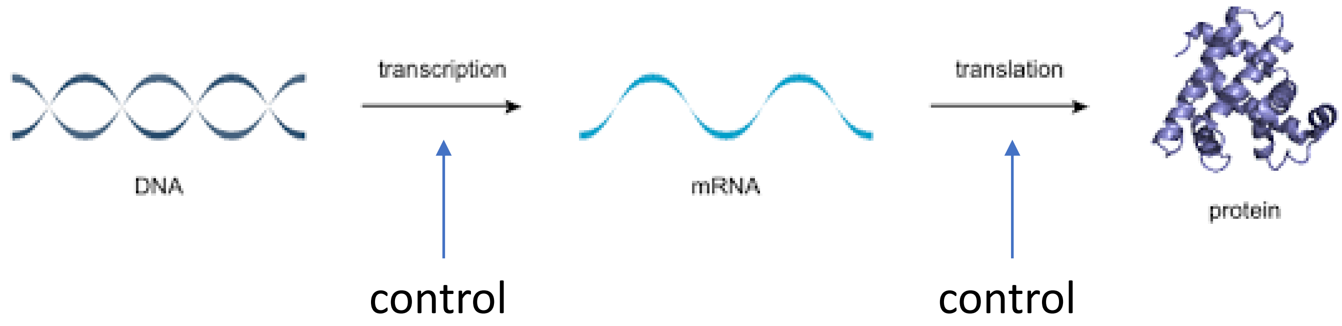
Can Network Stability promise Evolutionary Stability?

Robust Motifs are by Evolutionary Design



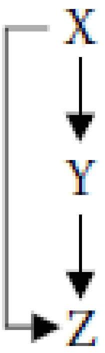
Some Gene X

Robust Motifs are by Evolutionary Design

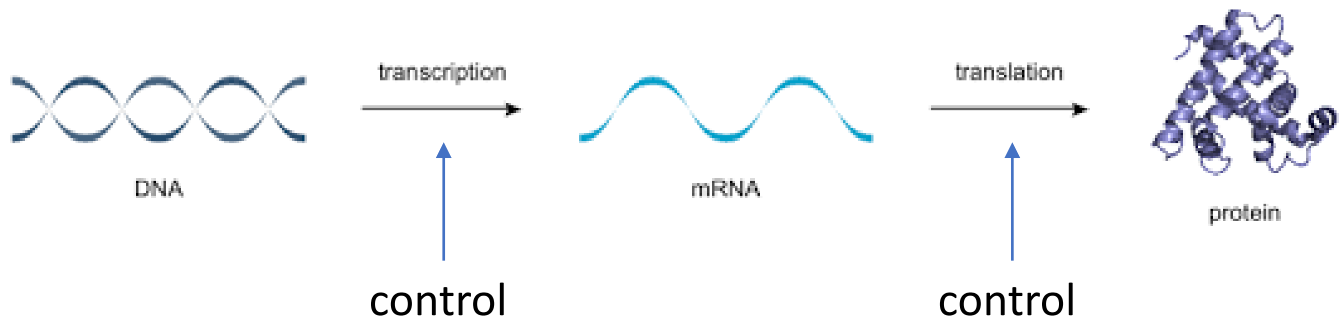


Some Gene X

Coherent
feed-forward loop
(C1-FFL)

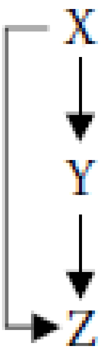


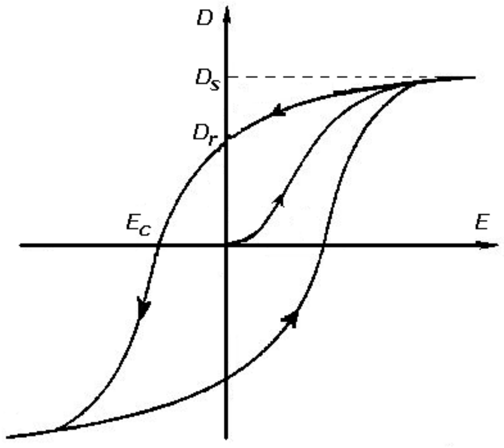
Robust Motifs are by Evolutionary Design



Some Gene X

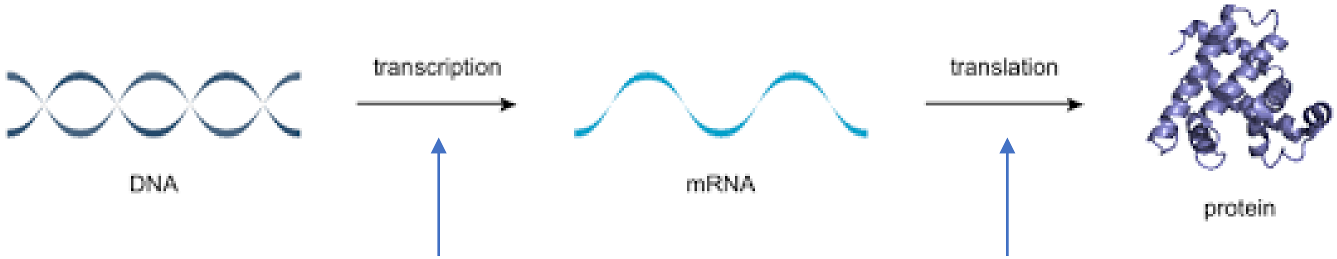
Coherent
feed-forward loop
(C1-FFL)





(Hysteresis)

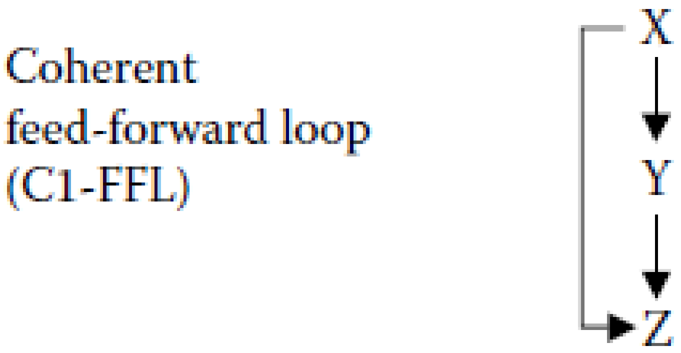
Robust Motifs are by Evolutionary Design

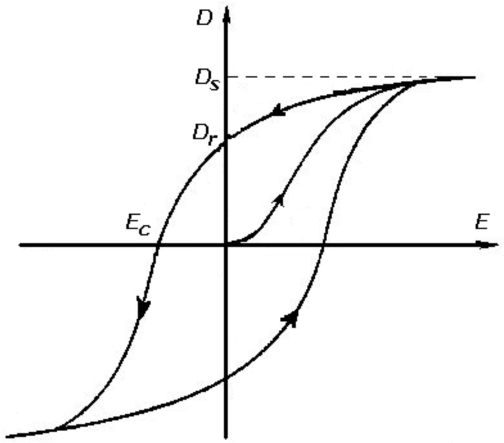


control

control

Some Gene X





(Hysteresis)

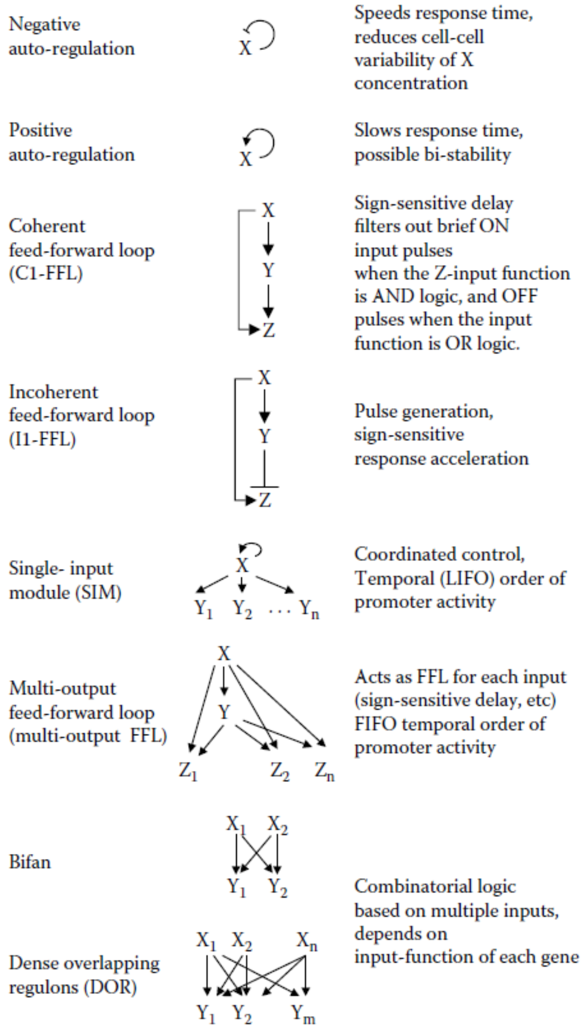


FIGURE 5.15 Network motifs in sensory transcription networks.

Robust Motifs are by Evolutionary Design

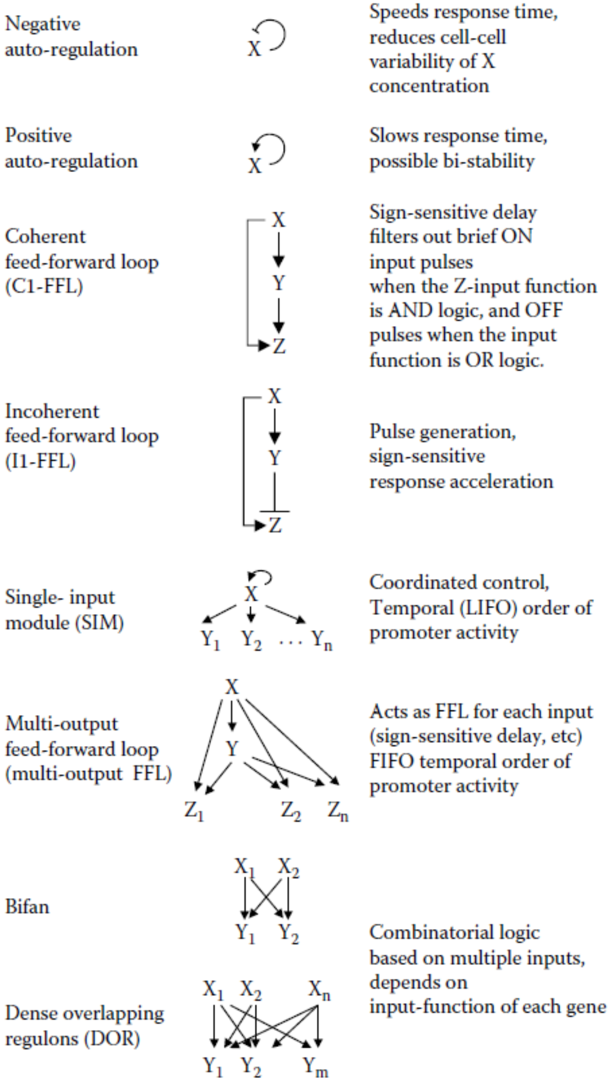
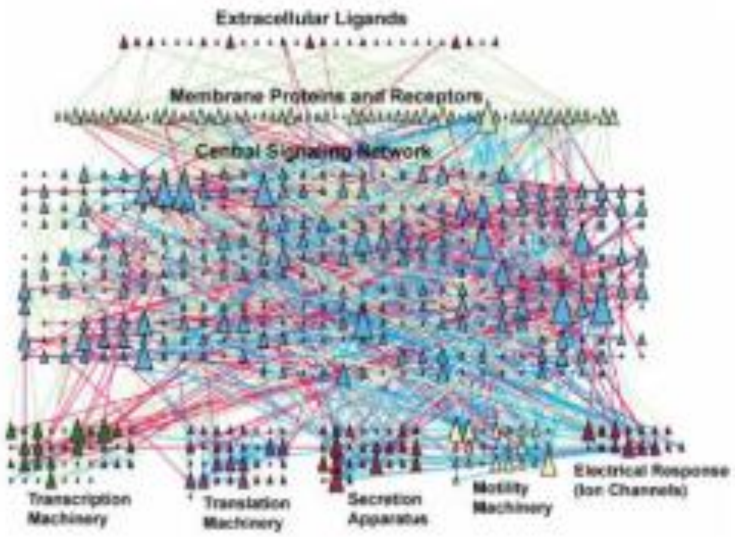
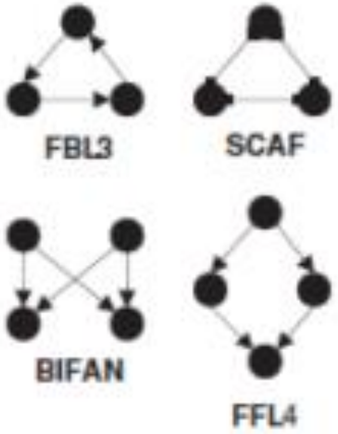


FIGURE 5.15 Network motifs in sensory transcription networks.



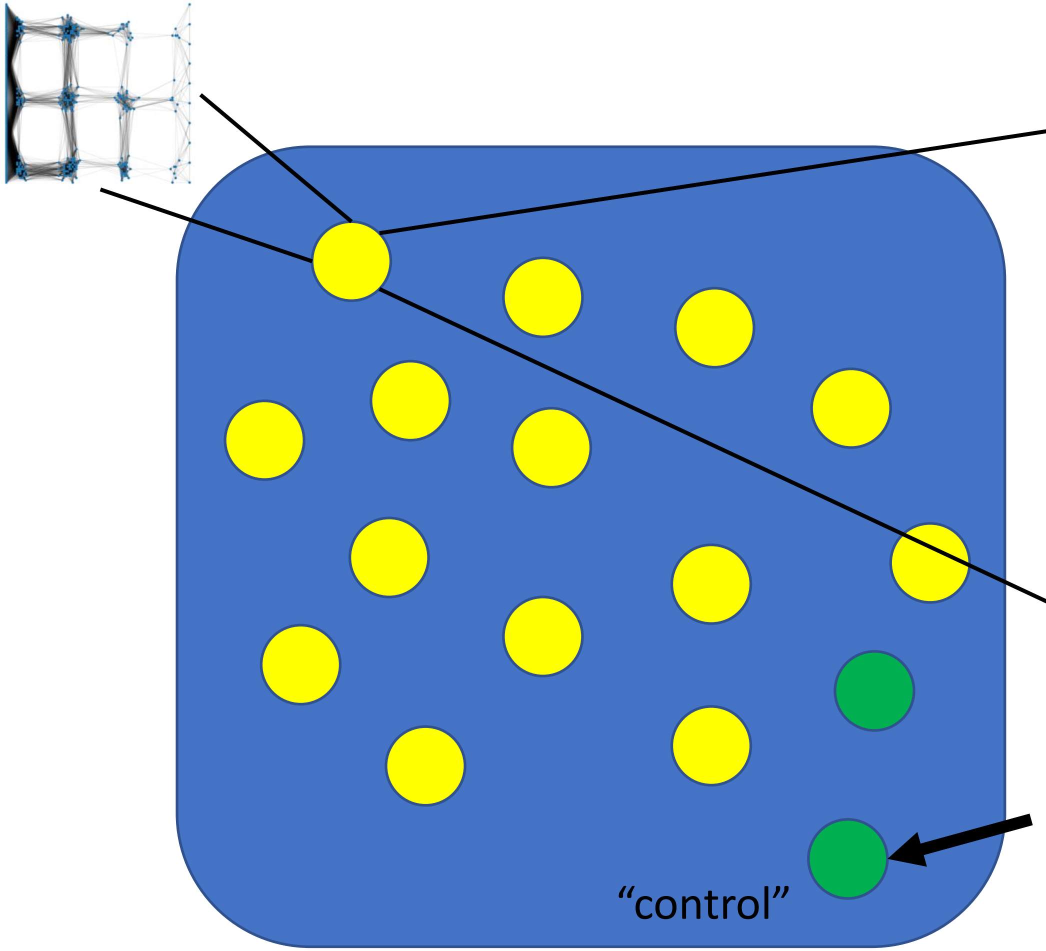


	Motifs counts		
Motif #	CN*	SN**	Z-score
FBL3	22	8.6 ± 8.1	4.34
SCAF	25	9.0 ± 3.4	4.72
BIFAN	1009	180.3 ± 28.0	29.60
FFL4	301	103.4 ± 17.2	11.51

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

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

Information Flow Stability informed Evolutionary Learning



B

$$
\begin{array}{c|ccccccccccccc}
 & \mathbf{I} & & & \mathbf{L1} & & & & & \mathbf{L2} & & & \mathbf{O} & \\
\hline
\mathbf{I} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\mathbf{L1} & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & 0 & 0 & 0 & \multicolumn{3}{c}{W_{L1(i,j)}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & 0 & 0 & 0 & & & & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\mathbf{L2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \multicolumn{2}{c}{W_{L2,i}} & 0 & 0 & 0 \\
\mathbf{O} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
\end{array}
$$

$$W$$

$$W - D$$

$$D = I$$

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

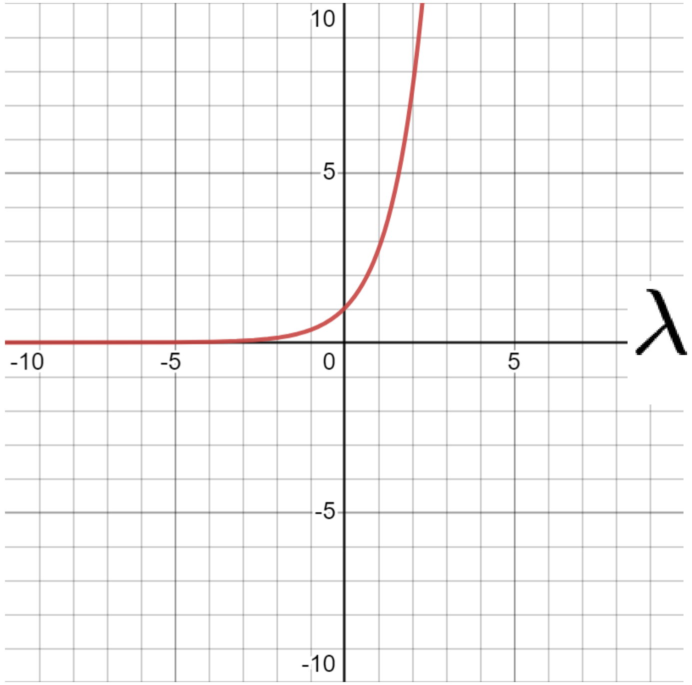
Cost function for Stability informed learning

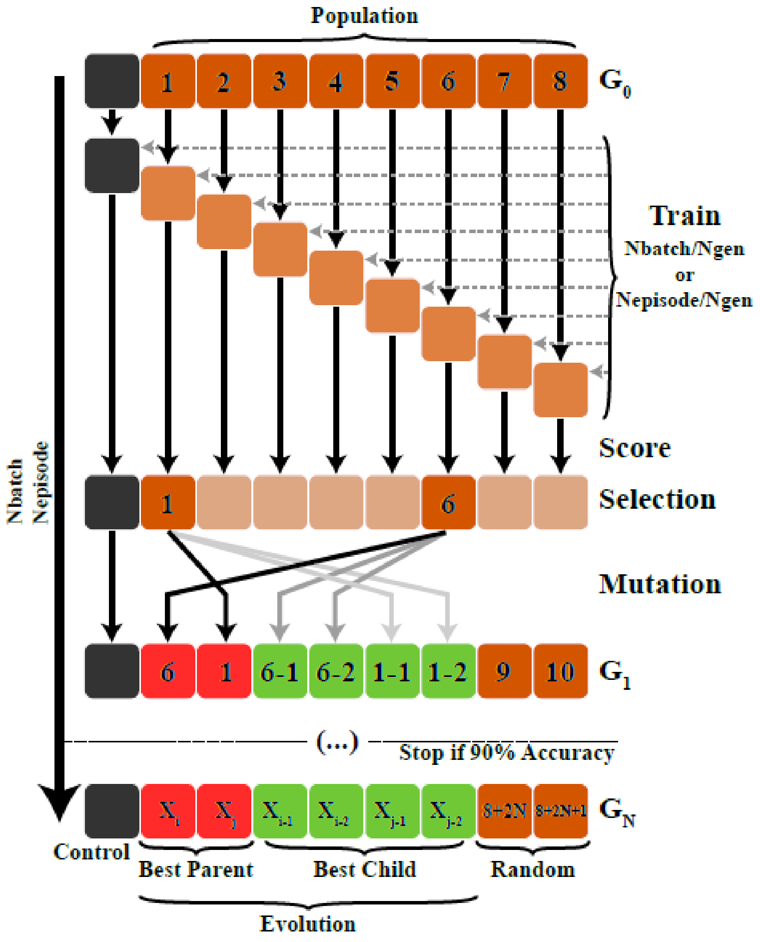
Stability term

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

Cross Entropy

Regularizer

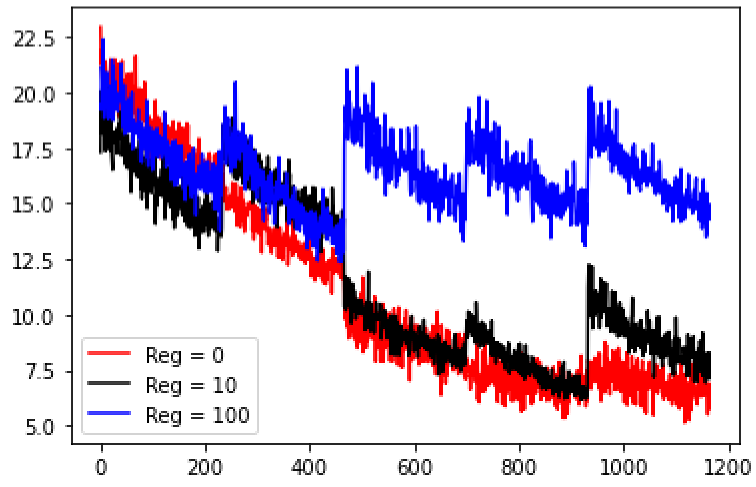




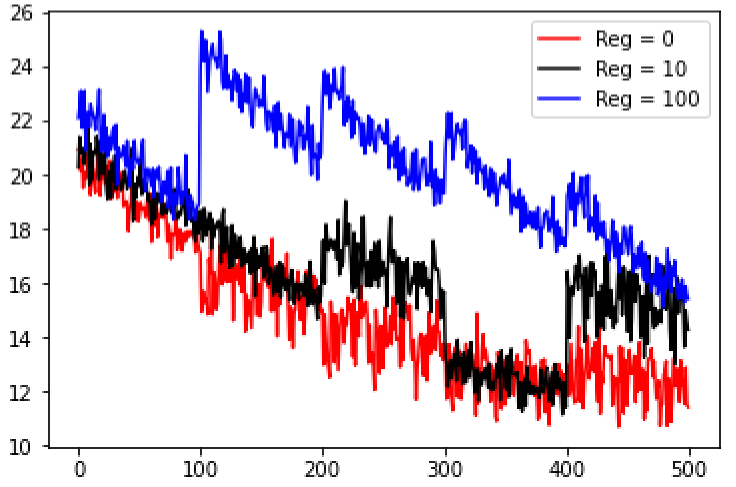
Structural convergence for stability informed evolution

CLASSIFICATION LOSS

BATCHE SIZE = 30



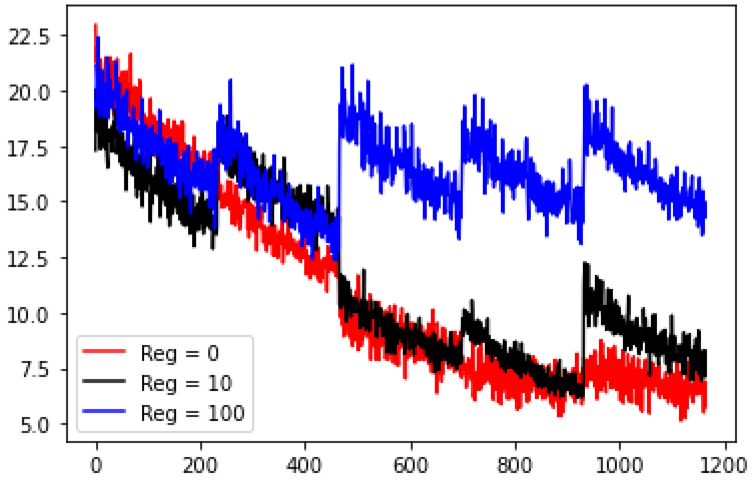
BATCHE SIZE = 40



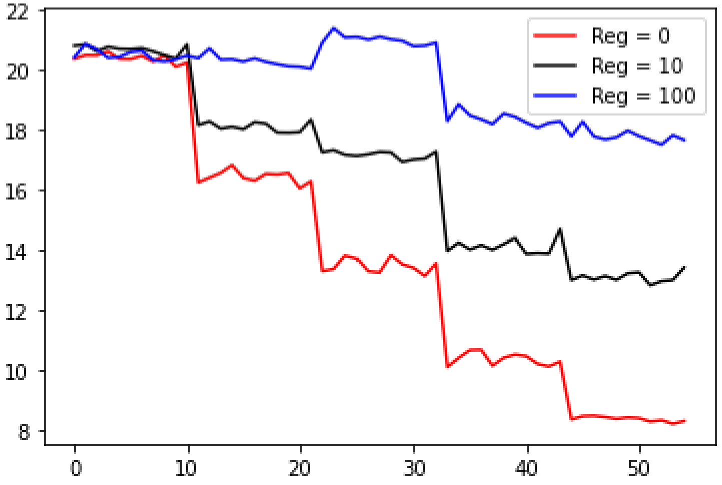
Structural convergence for stability informed evolution

CLASSIFICATION LOSS

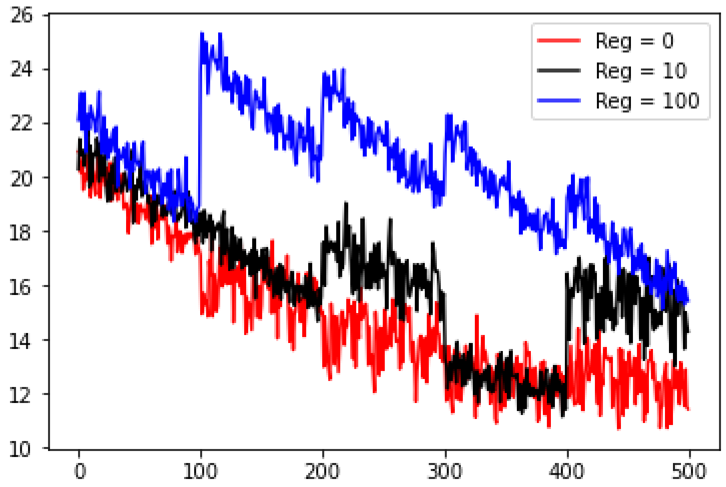
BATCHE SIZE = 30



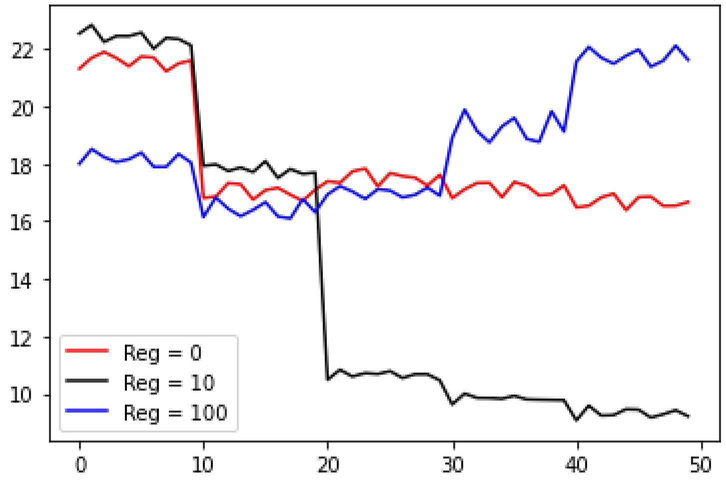
BATCHE SIZE = 500



BATCHE SIZE = 40



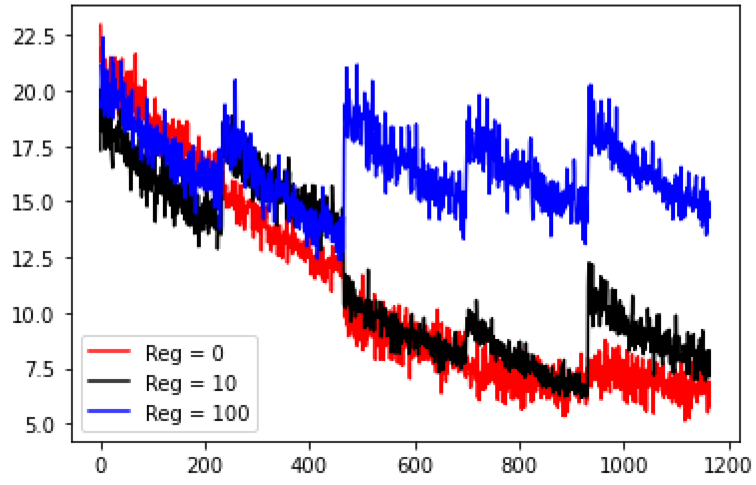
BATCHE SIZE = 700



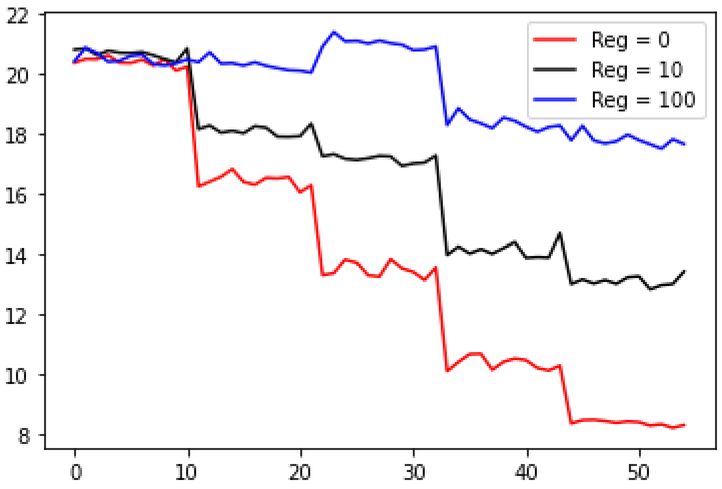
Stability informed evolution outperformed SGD for a narrow batch size

CLASSIFICATION LOSS

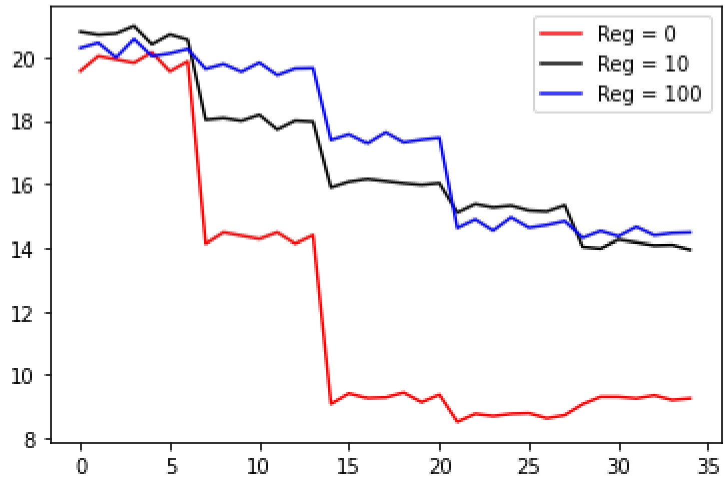
BATCHE SIZE = 30



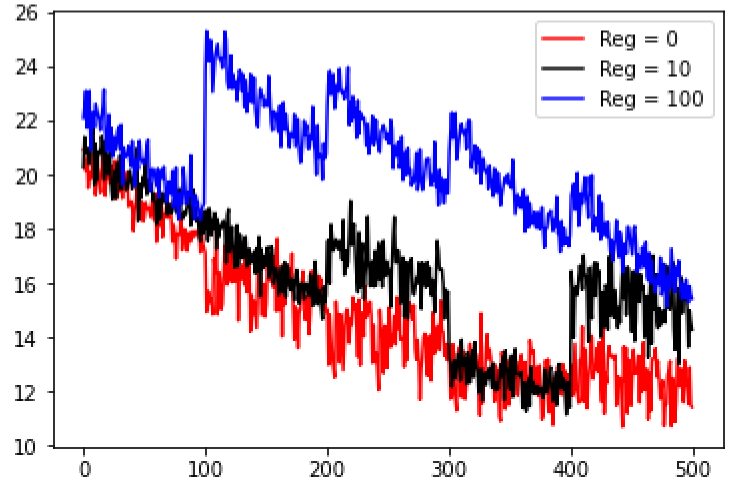
BATCHE SIZE = 500



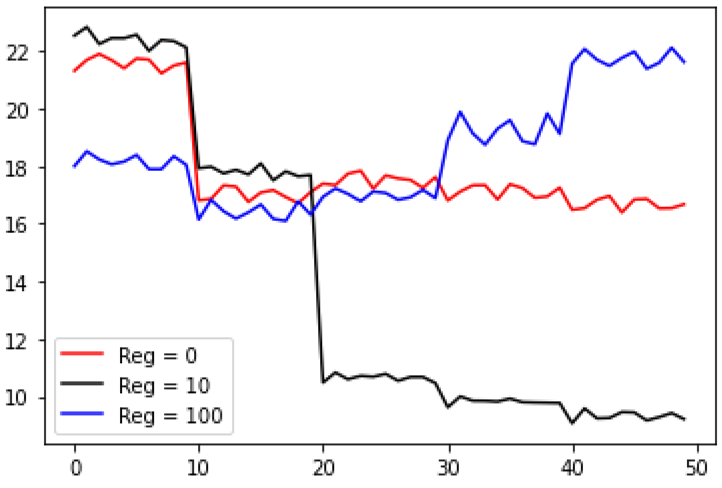
BATCHE SIZE = 1000



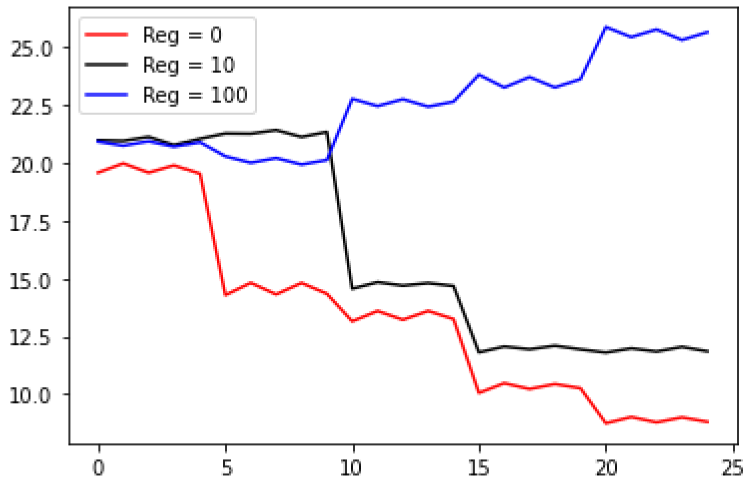
BATCHE SIZE = 40



BATCHE SIZE = 700



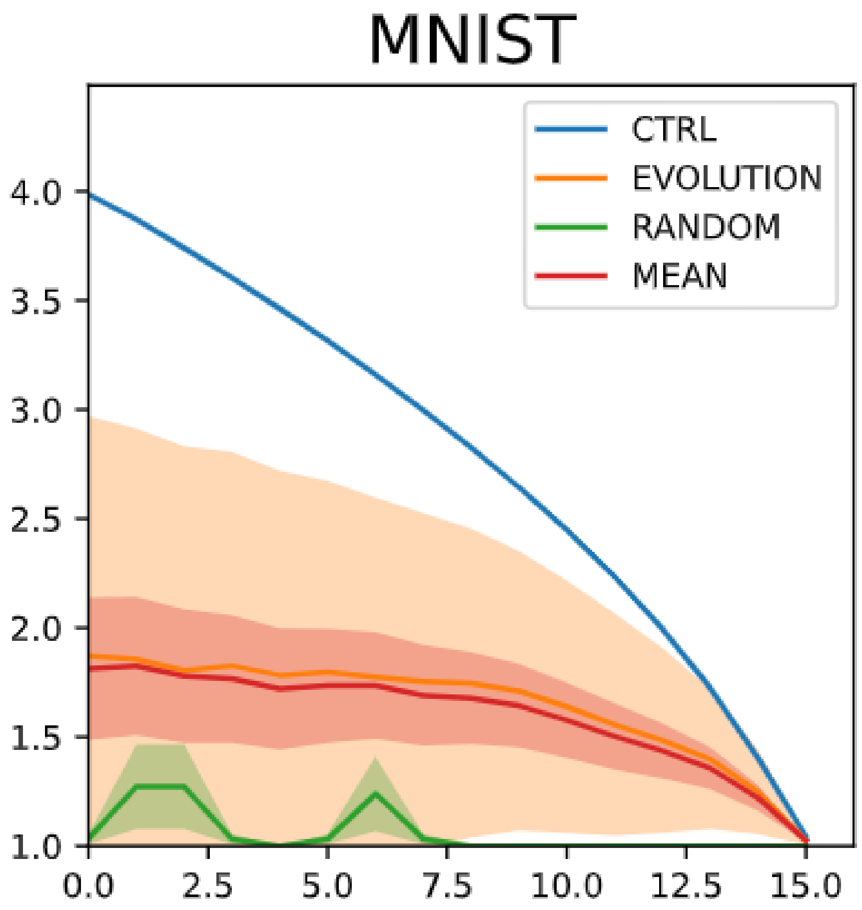
BATCHE SIZE = 1300



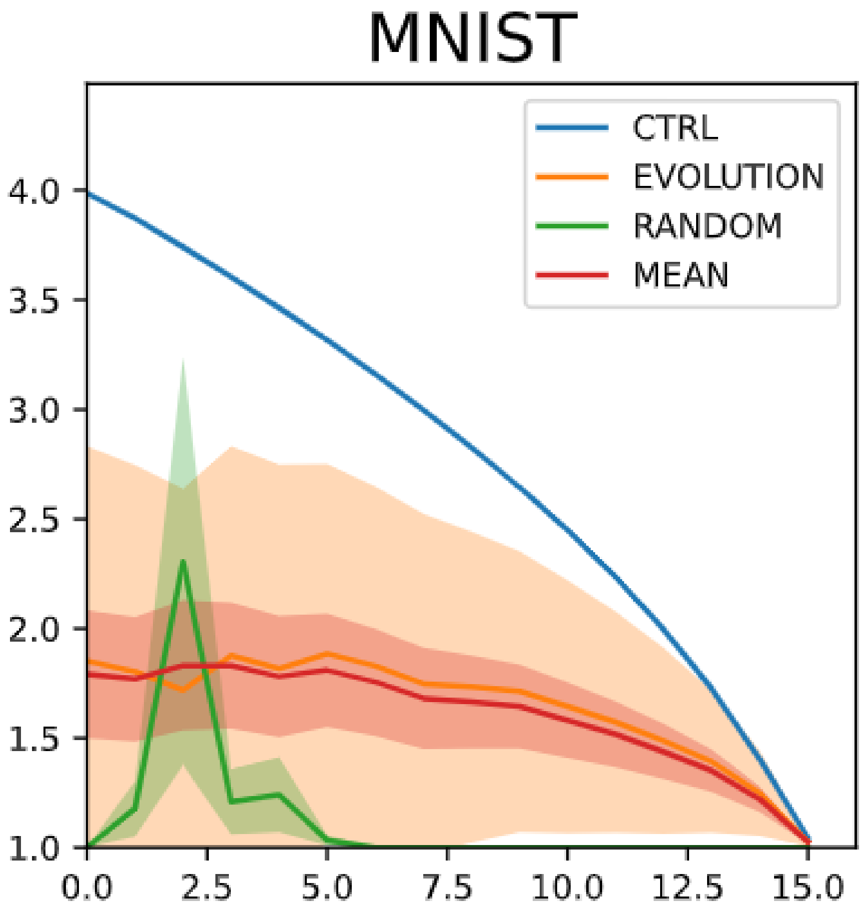
BATCHES

Evolutionary Stability is not enhanced by Network Stability

Reg = 0



Reg = 1



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!