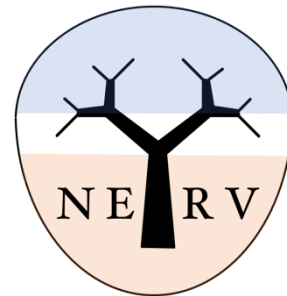


Emergence of Giant Component in Percolated Artificial Neural Networks

Adrien Pontlevy and Fabrizio, de Vico Fallani



Inria

 Paris Brain
Institute

 cnrs

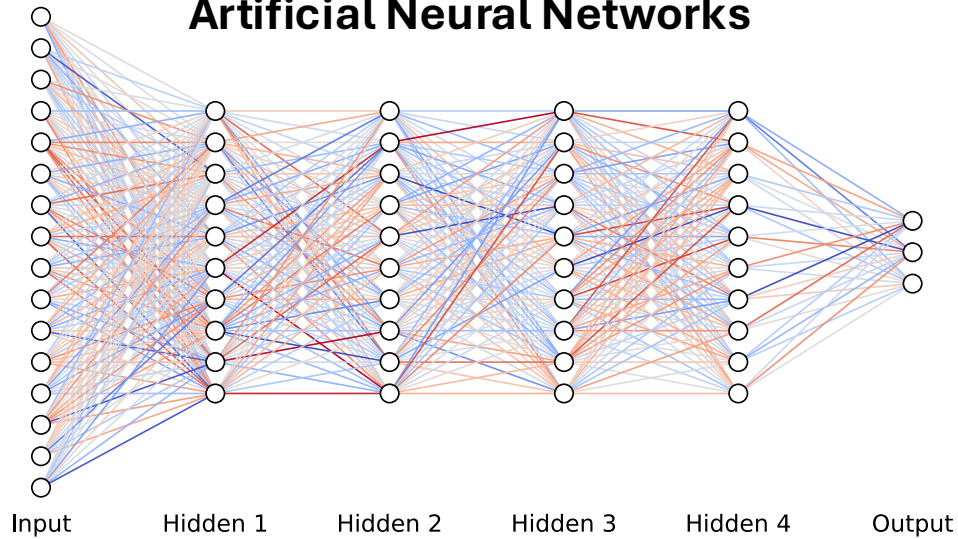
 SORBONNE
UNIVERSITÉ

 Hôpital
Pitié-Salpêtrière
AP-HP

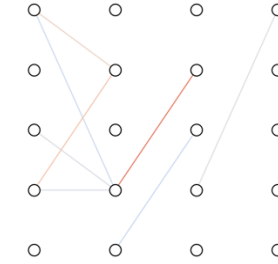
 Inserm

Context

Artificial Neural Networks

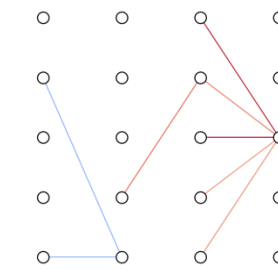
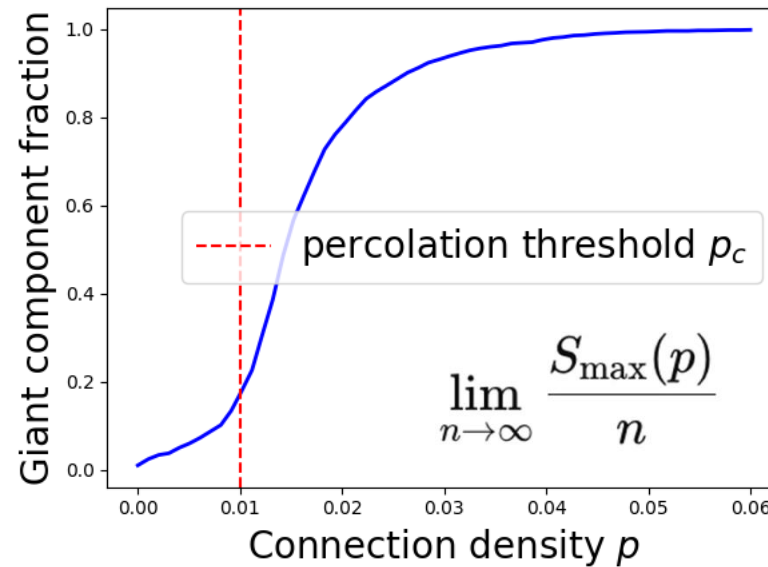
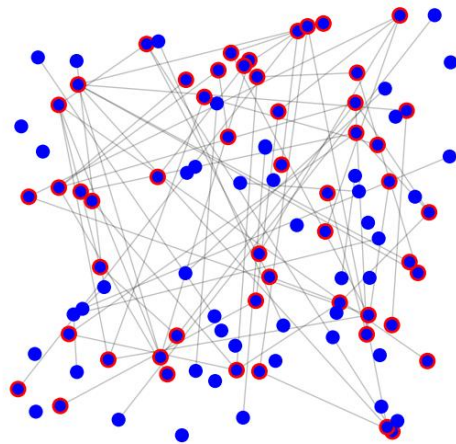


Pruning strategies



Random pruning

Percolation

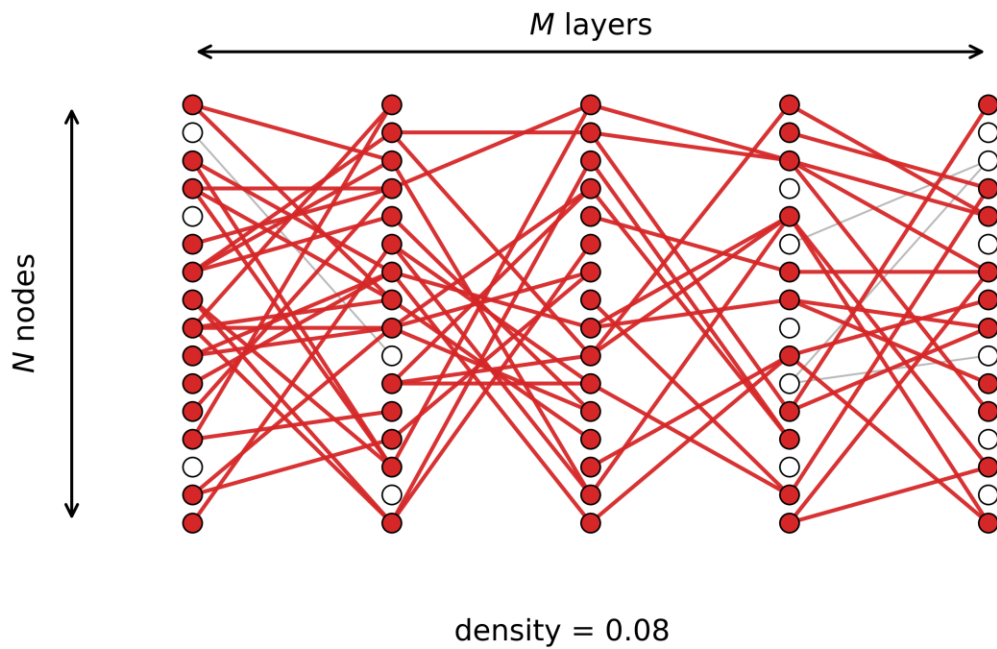


Magnitude-based pruning

Questions

1. What is the percolation threshold in Multi-Layer Perceptrons (MLPs) for Bernoulli percolation ?
2. How structural and functional robustness relate under edge removal ?

Bernoulli Percolation in Layered Networks

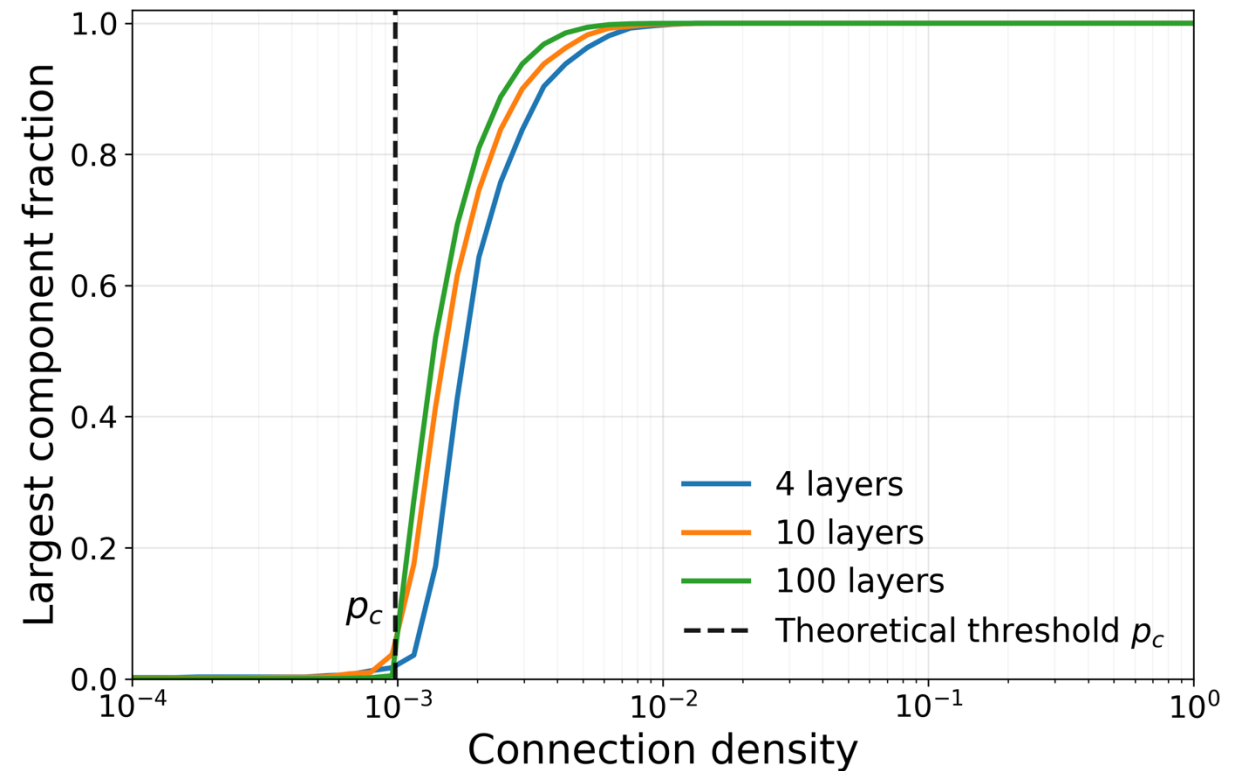


For Multi-layer perceptrons,

$$\langle k \rangle \simeq 2N$$

\Rightarrow

$$p_c \simeq \frac{1}{2N - 1}$$

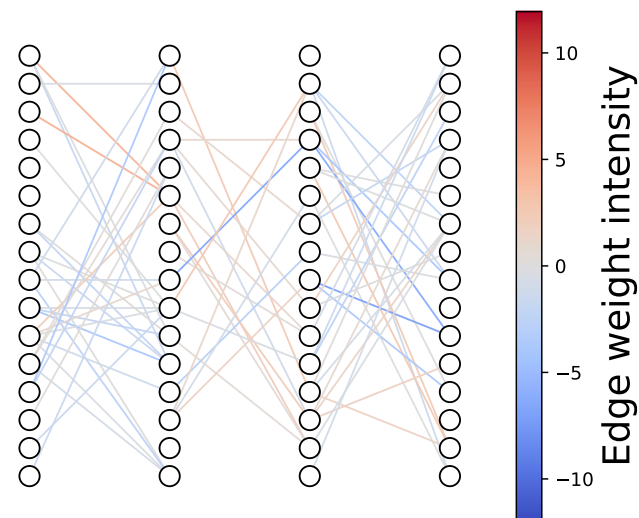
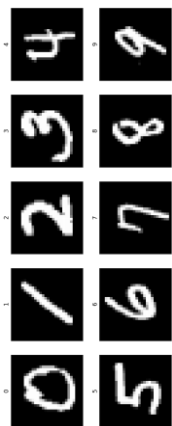


Percolation threshold

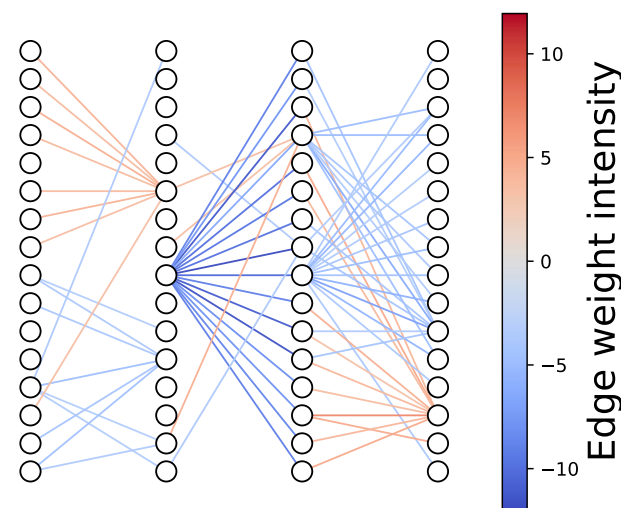
$$p_c = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}$$

Cohen et al., PRL 2000

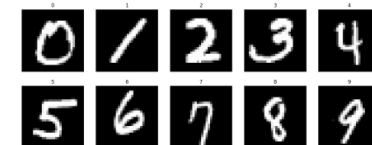
Comparing Bernoulli and Descending Percolation



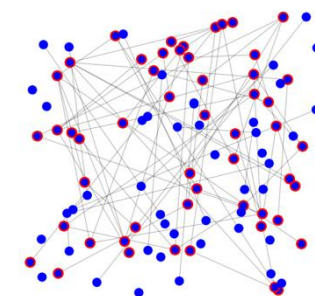
Bernoulli percolation



Descending percolation

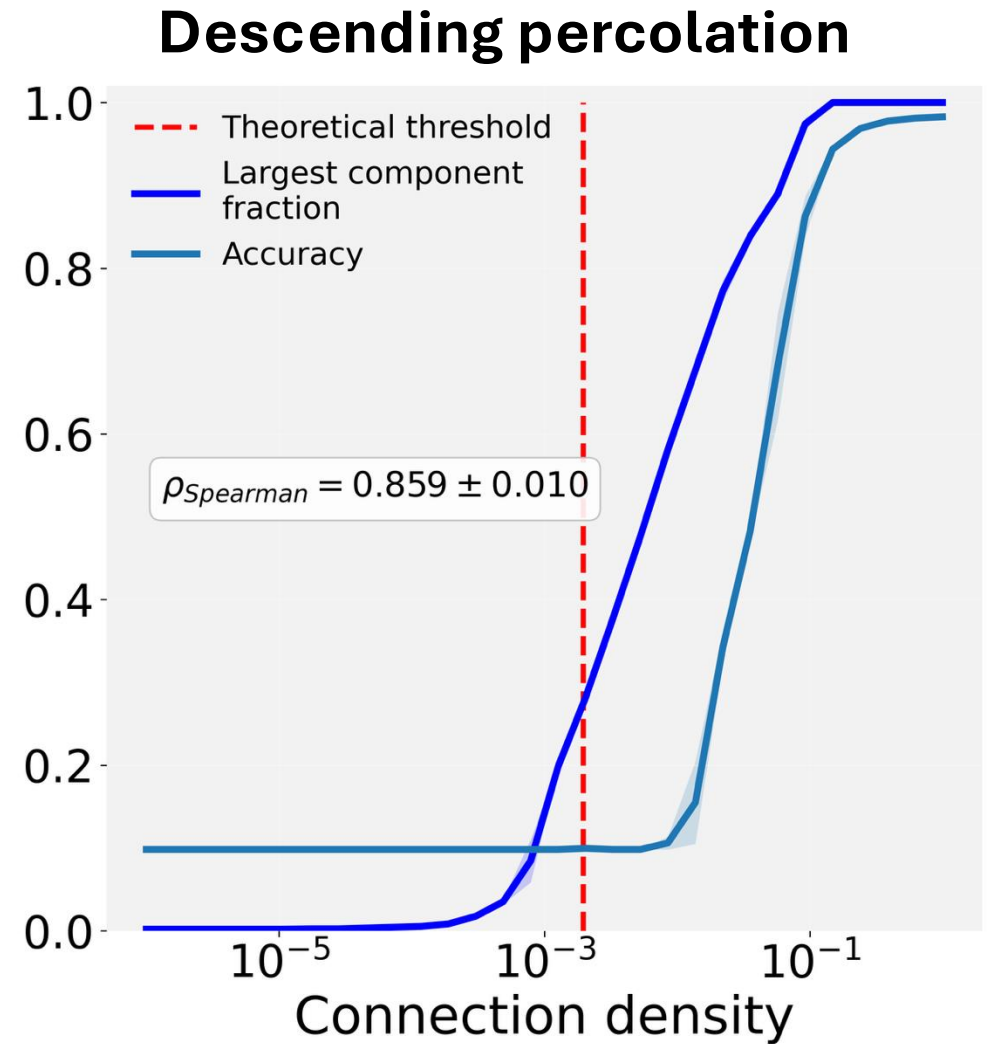
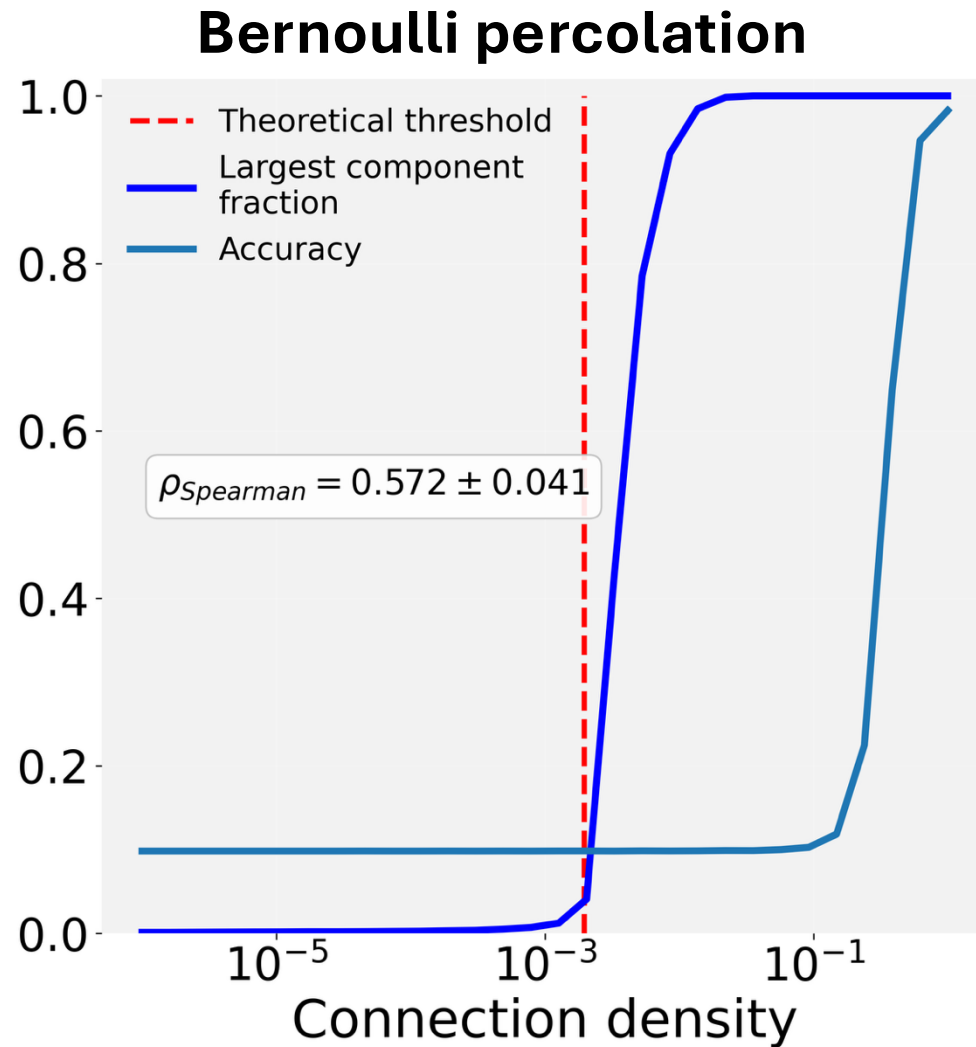


Accuracy ?



Largest Component fraction ?

High-magnitude weights are structurally organized



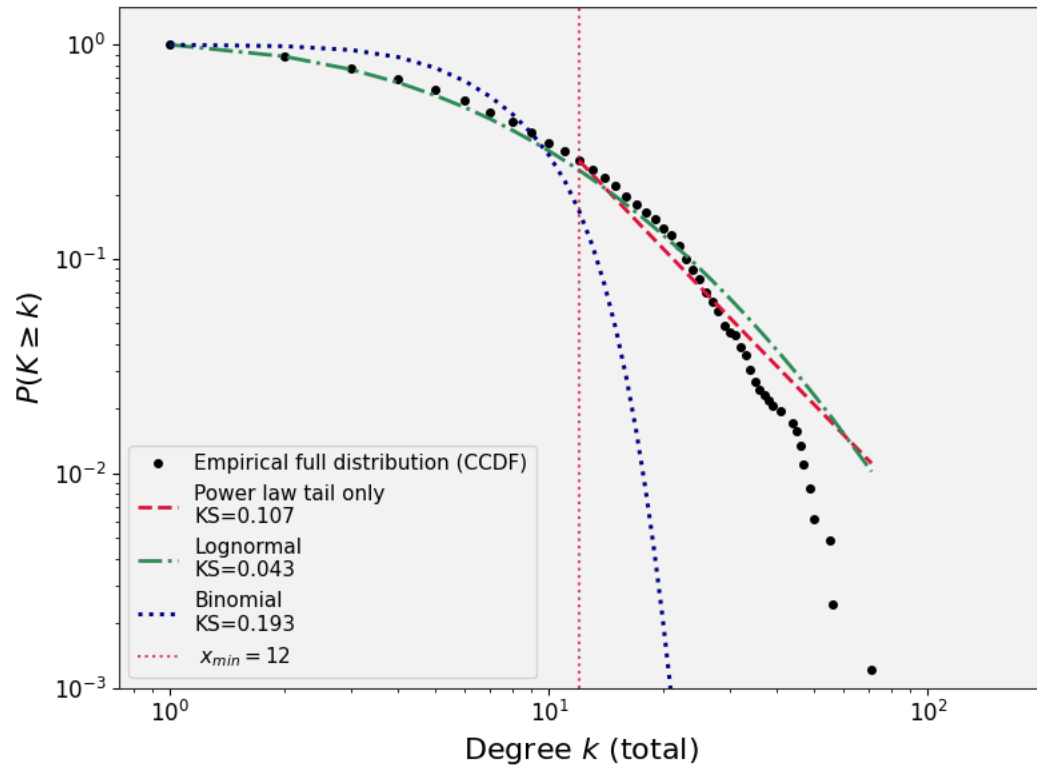
This comparison reveals a non-trivial organization of the high-magnitude weights in the network !

Additional question

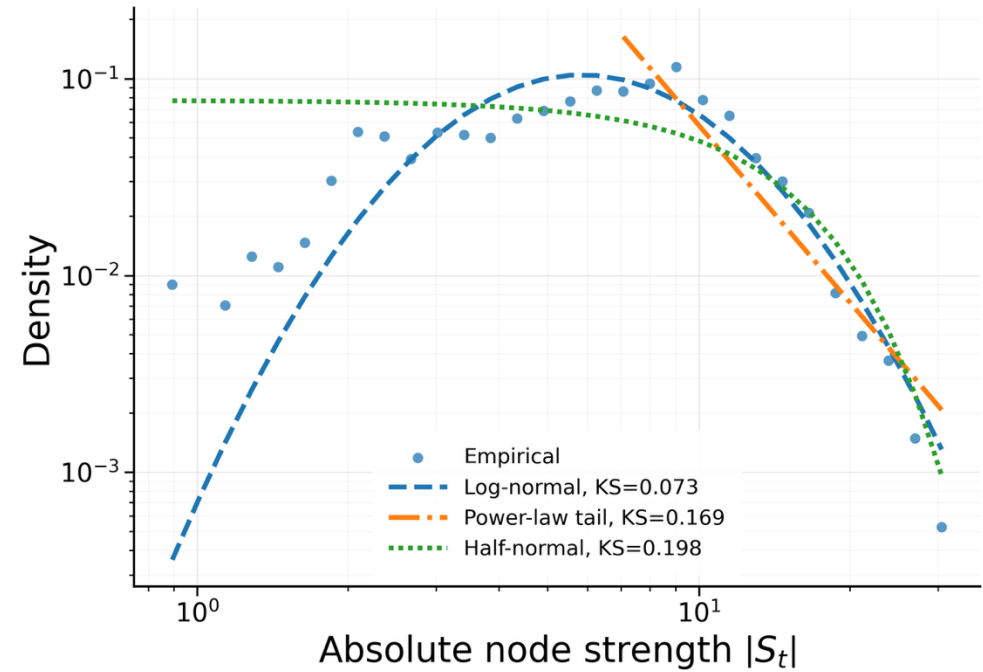
How is weight mass distributed across nodes and does it follow a known statistical law?

Degree/Strength Distributions are heavy-tailed

Degree distribution of descending-pruned networks



$$S_i = \sum_j w_{ij} \longrightarrow |S_i| \longrightarrow P(|S|)$$



Pontlevy & de Vico Fallani, Ready to submit

Conclusion

What we found ?

- Bernoulli percolation threshold in deep networks is layer-number independent
- The weight magnitude of the weights provide function and structure information, and the training process leaves a detectable structural fingerprint
- The precocous emergence of a giant component is reflected by a heavy-tailed structure.

Perspectives?

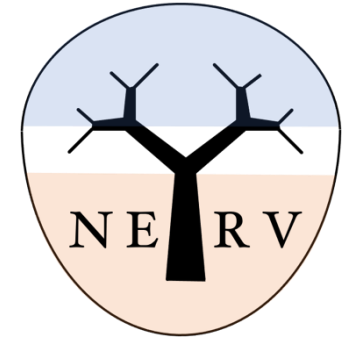
- Objective-based pruning

$$s_i = \frac{1}{2} \frac{\partial^2 \mathcal{L}}{\partial w_i^2} \cdot w_i^2$$

LeCun et al., NeurIPS 1989

- Other architectures ?

Acknowledgements



Thank you for your attention!