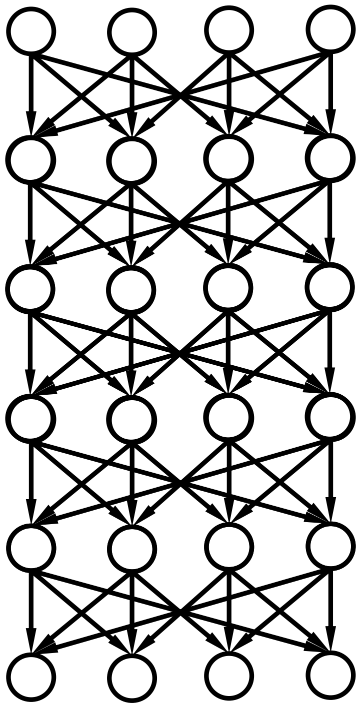
Graph Structure of Neural Networks

Jiaxuan You[1], Jure Leskovec[1], Kaiming He[2], Saining Xie[2]

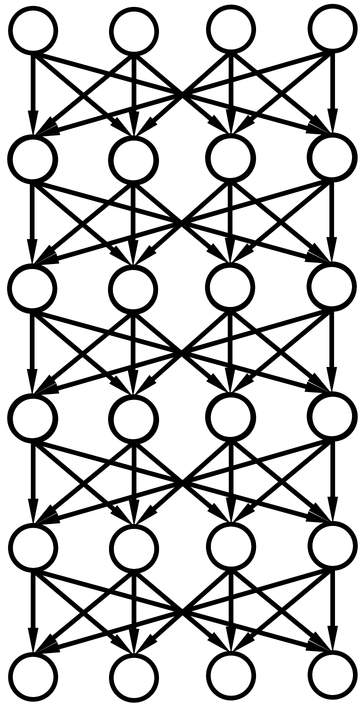
Stanford University[1]
Facebook AI Research[2]

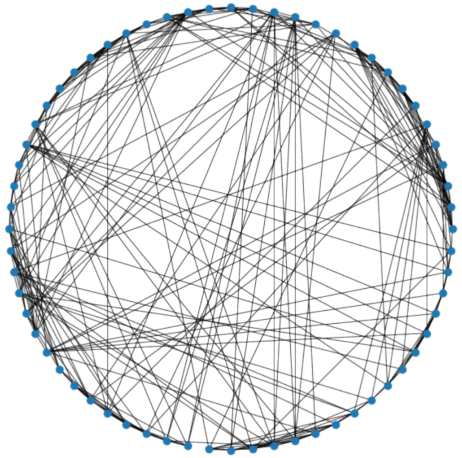


Neural networks consist of neurons

We know connections between neurons affect NN performance

But how?





Underneath a NN, there is a graph

We want to find a proper **graph representation of NN** to answer:

*Is there a link between the **graph structure** and **NN performance**?*

*If so, what are **structural signatures** of well-performing NNs?*

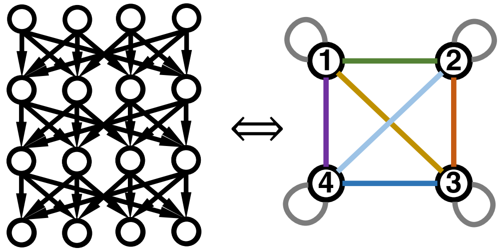
*Can these signatures **generalize** across tasks and datasets?*

…

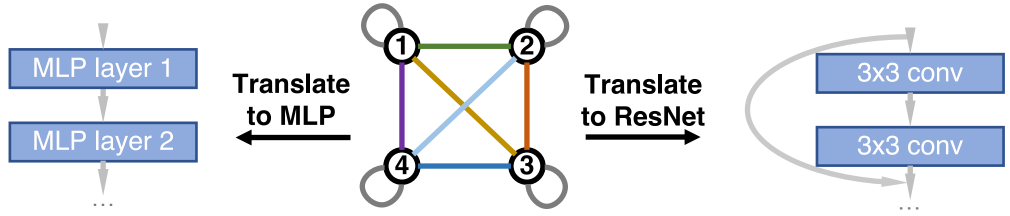
J. You, J. Leskovec, K. He, S. Xie, *Graph Structure of Neural Networks, ICML 2020*

Overview: **Methodology**

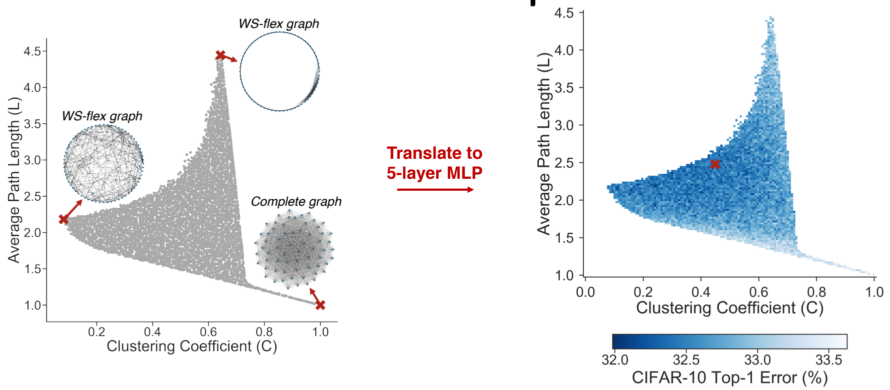
- A novel representation of neural networks: relational graphs



- Relational graphs can represent diverse neural architectures

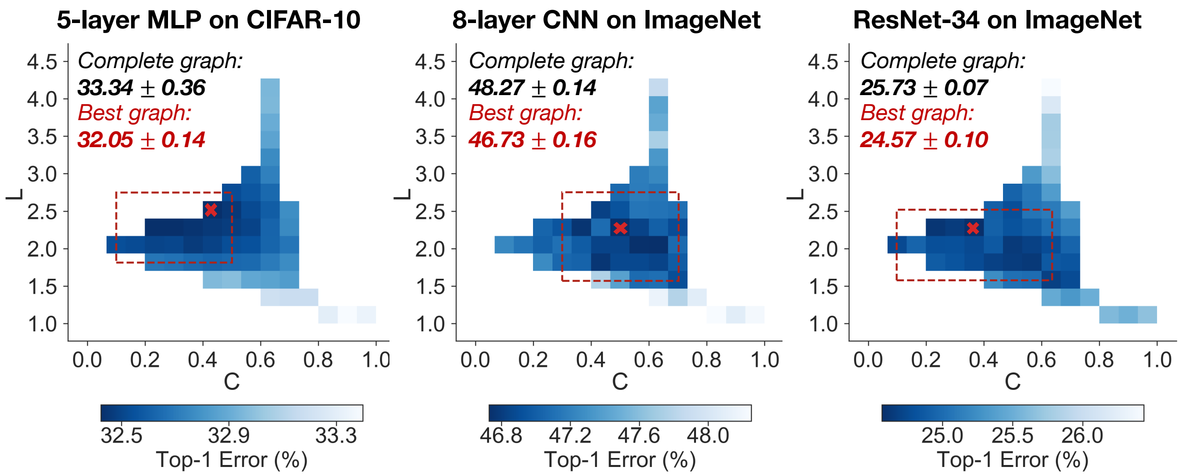


- Tools from network science → Graph structure *vs* NN performance



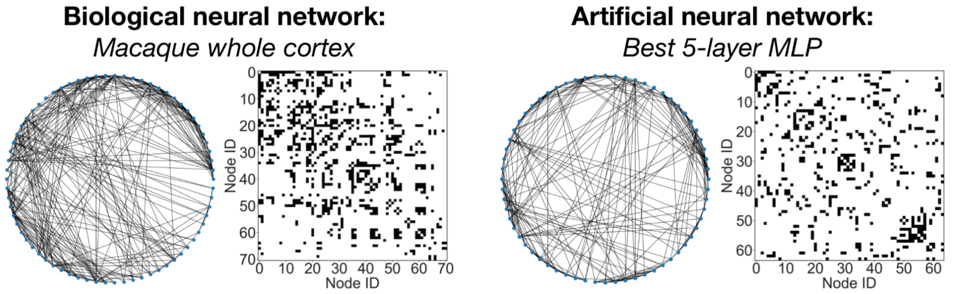
Overview: Key Findings

- Consistent “Sweet Spot” for top NNs across architectures



  - Graphs with certain structure measures consistently performs well (controlling computational budgets)

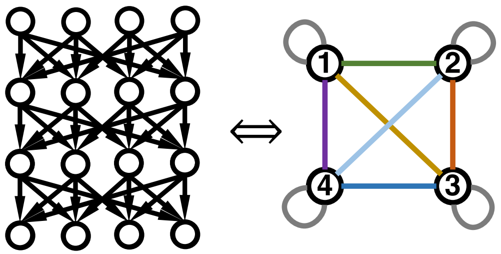
- Top artificial NNs are similar to real biological NNs



  - Graph structure of the best 5-layer MLP we found, is similar to the macaque whole cortex network

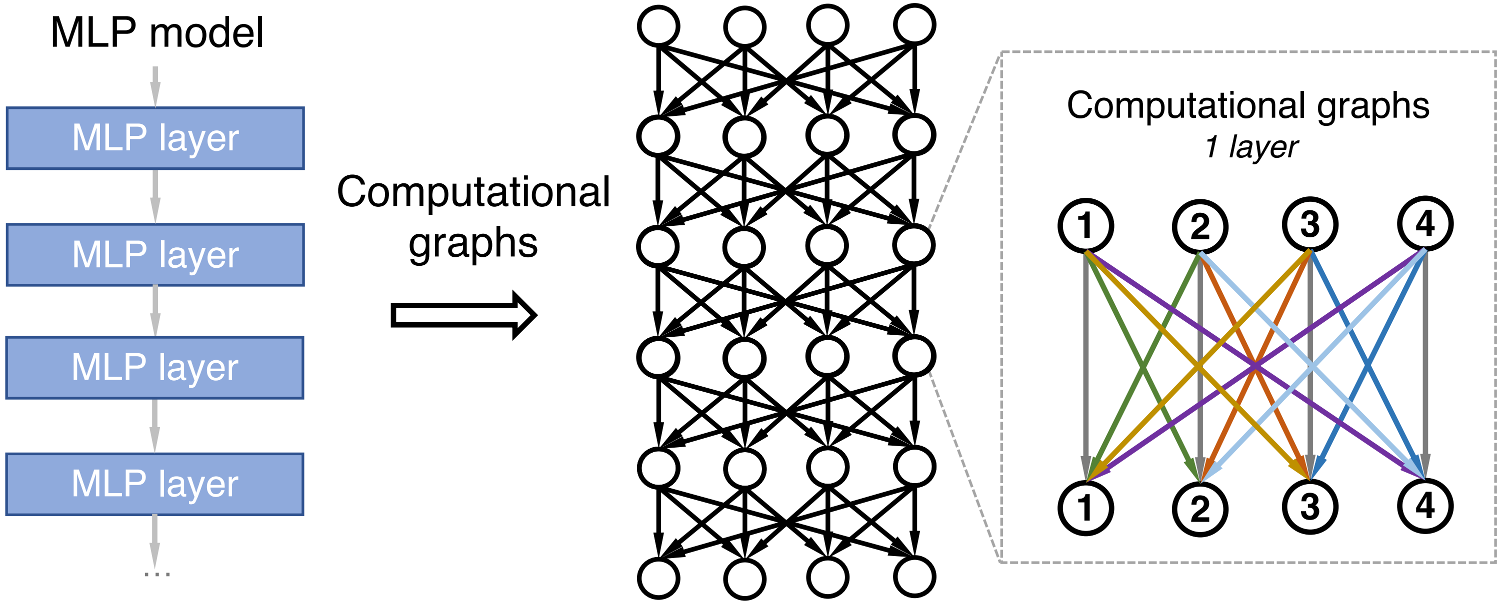
Overview: **Methodology**

- A novel representation of neural networks: relational graphs ←
  - Computational vs. Relational graphs
  - Neural computation as message exchange on relational graphs



- Relational graphs can represent diverse neural architectures
- Tools from network science → Graph structure vs NN performance

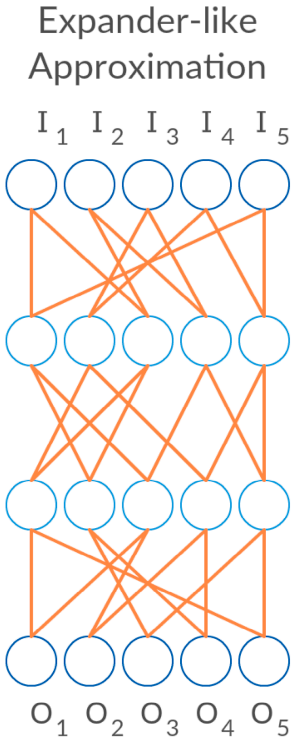
Background: NNs as Computational Graphs



Background: Related Work

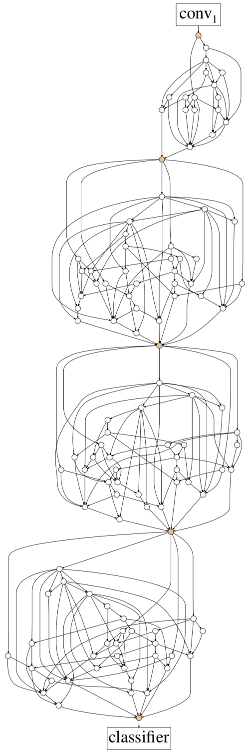
Existing graph-based architecture design approaches focus on computational graphs

Deep Expander Networks, Prabhu et al., 2018



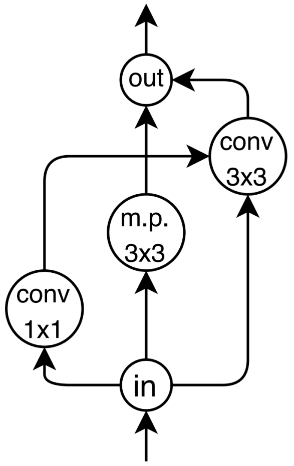
Generate bipartite graphs over neurons

RandWire, Xie et al., 2019



Generate directed acyclic graphs over NN layers

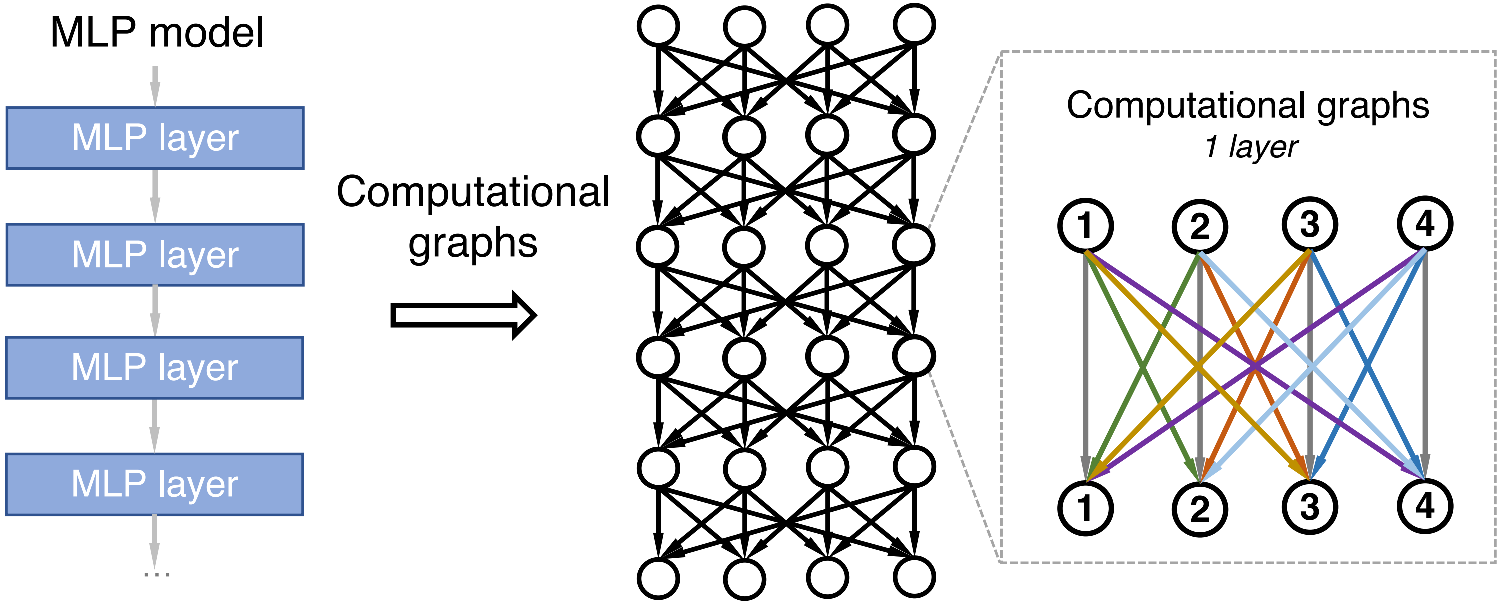
NAS-Bench-101, Ying et al., 2019



$$\begin{pmatrix} 0, 1, 1, 1, 0 \\ 0, 0, 0, 1, 0 \\ 0, 0, 0, 0, 1 \\ 0, 0, 0, 0, 1 \\ 0, 0, 0, 0, 0 \end{pmatrix}$$

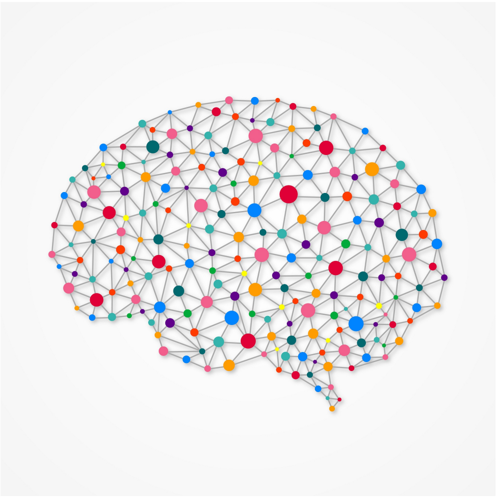
[in, conv1x1, mp3x3, conv3x3, out]

Limitations: NNs as Computational Graphs



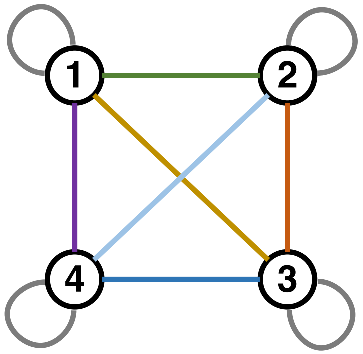
Limitations:

- Lack of flexibility
  - Directed acyclic graphs
- Disconnection with neuroscience
  - Brain networks have flexible structure
  - Bi-directional information exchange



Our Approach: **Relational Graphs**

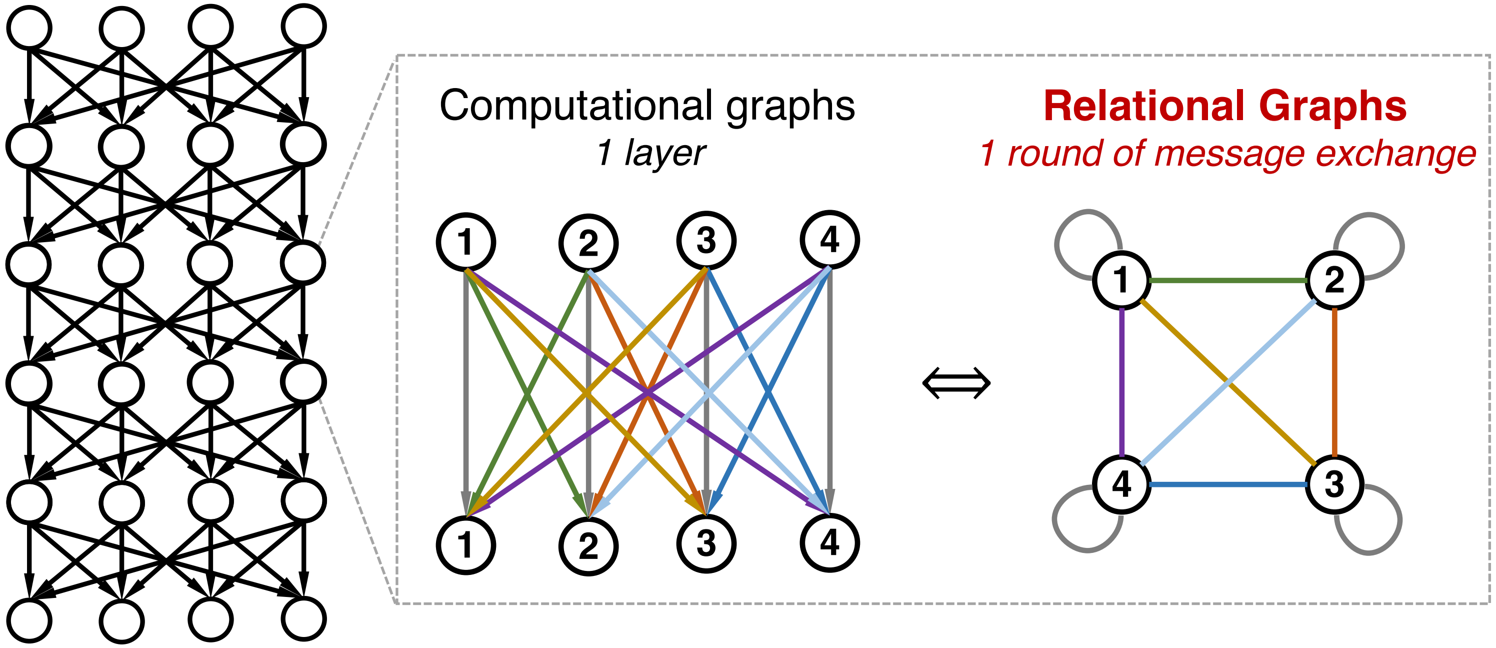
Relational Graphs
1 round of message exchange



Relational graph definition:

- **Nodes** are neurons
- **Edges** specify (undirected) connectivity between neurons;
- **Computation** is conducted by message exchange over the graph structure, where a node exchange messages with its neighbors

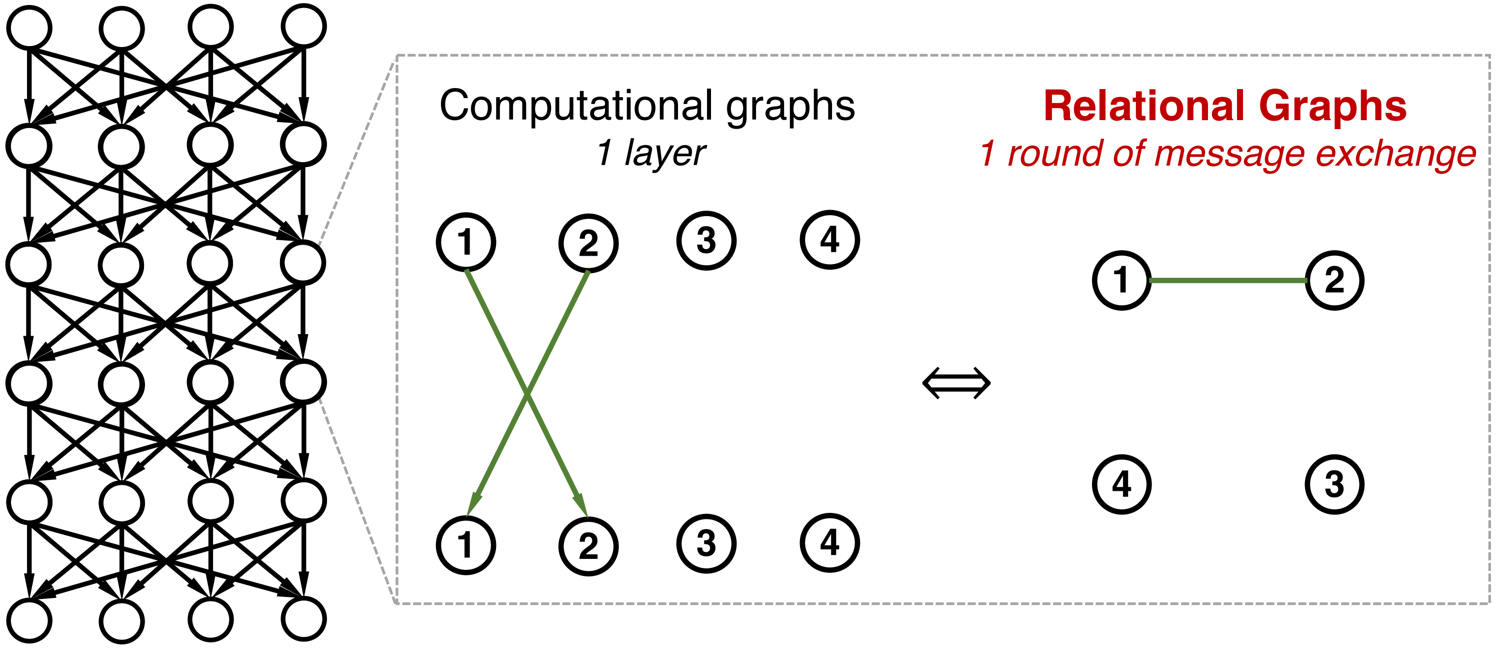
Our approach: Relational Graphs



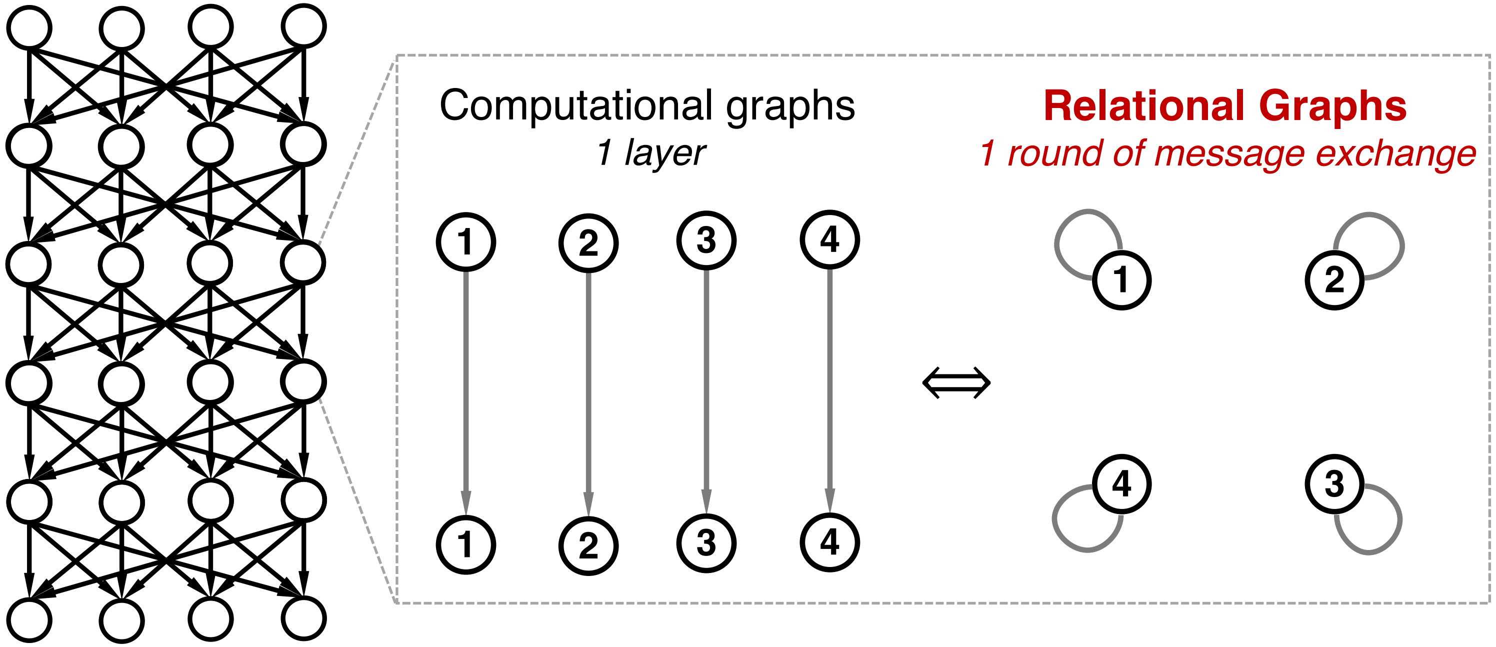
Relational graph definition:

- **Nodes** are neurons
- **Edges** specify (undirected) connectivity between neurons;
- **Computation** is conducted by message exchange over the graph structure, where a node exchange messages with its neighbors

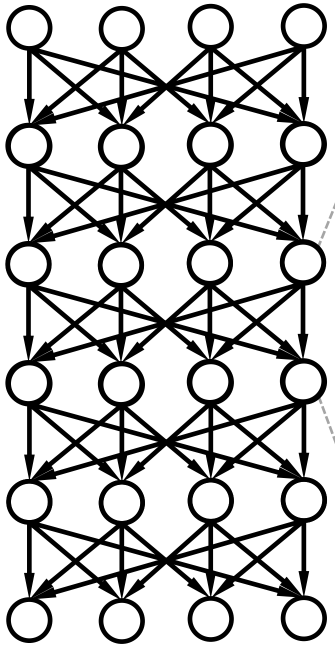
Our approach: Relational Graphs

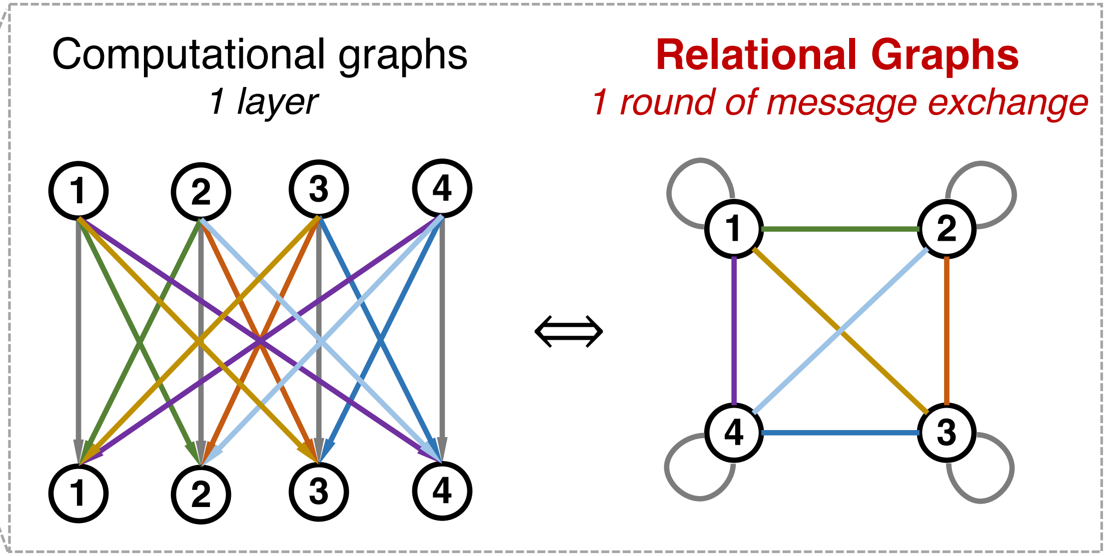


Our Approach: Relational Graphs



Benefits of Relational Graphs

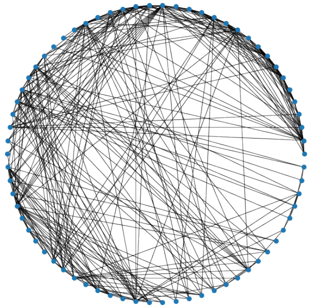




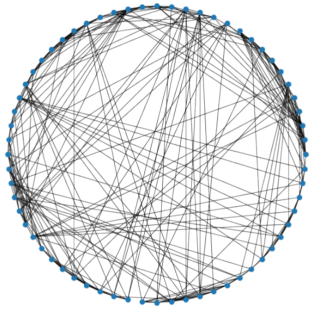
Benefits:

- Flexibility
  - No restrictions on graph structure (we focus on undirected graph)
- Connections with neuroscience
  - Bi-directional information exchange

Biological neural network:
Macaque whole cortex

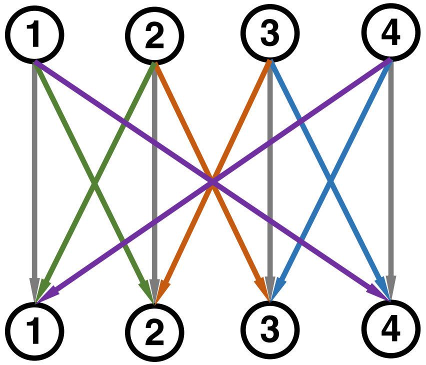


Artificial neural network:
Best 5-layer MLP



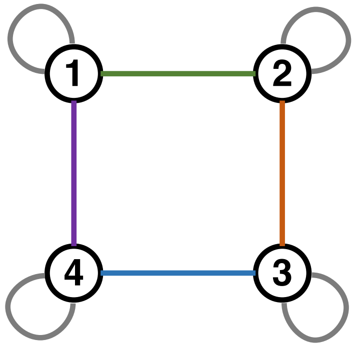
Diverse Relational Graphs

Computational graphs
1 layer

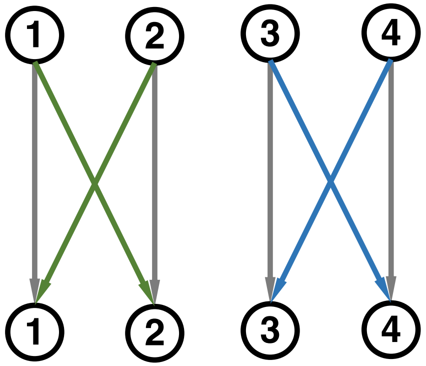


⟺

Relational Graphs
1 round of message exchange

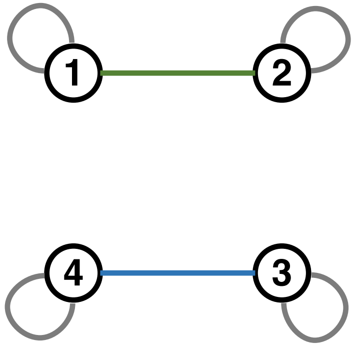


Computational graphs
1 layer

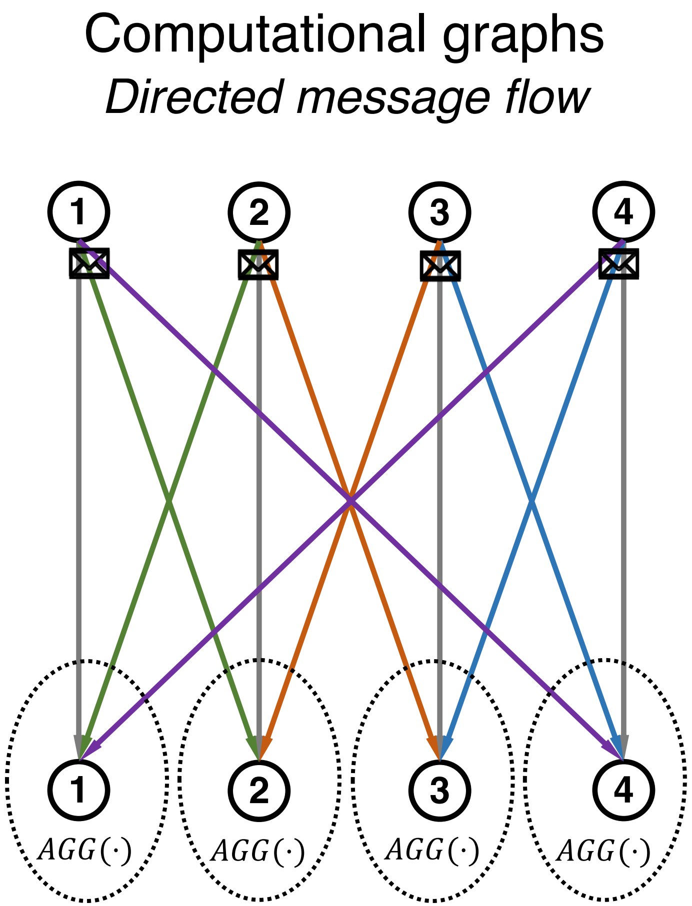


⟺

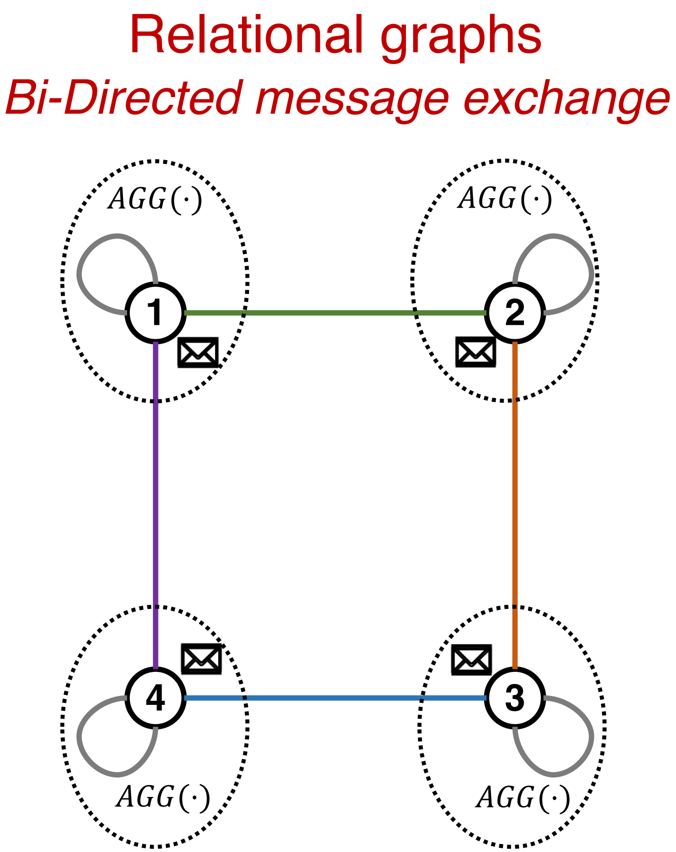
Relational Graphs
1 round of message exchange



Neural Computation as **Message Exchange**

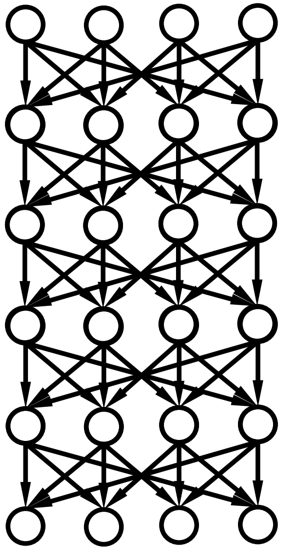


$\Longleftrightarrow$



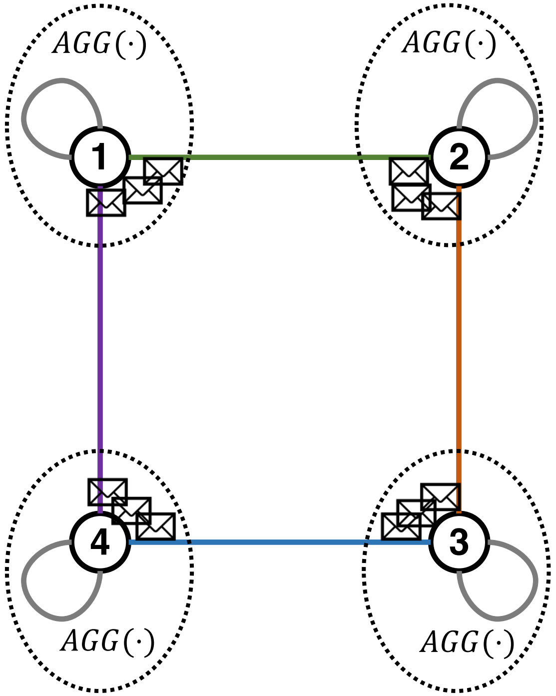
Neural Network as **Rounds of Message Exchange**

A 5-layer
Neural network



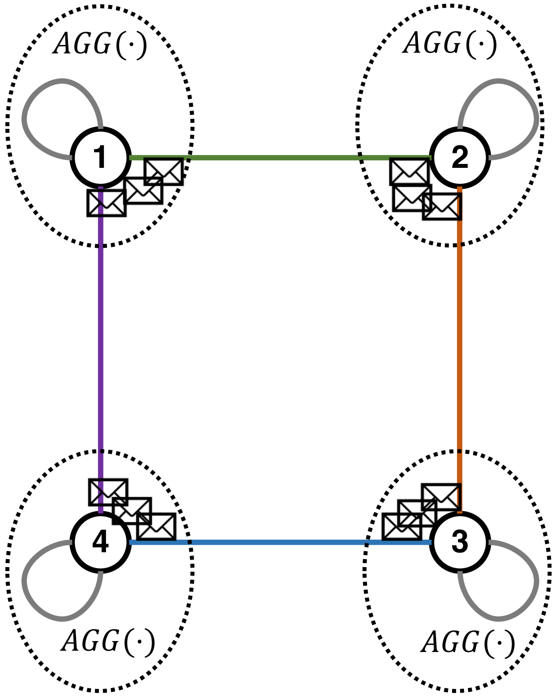
$\Longleftrightarrow$

A relational graph with
5 rounds of message exchange



This is how Graph Neural Networks
compute embeddings!

Side Note: Connections with **GNNs**



This is how Graph Neural Networks compute embeddings!

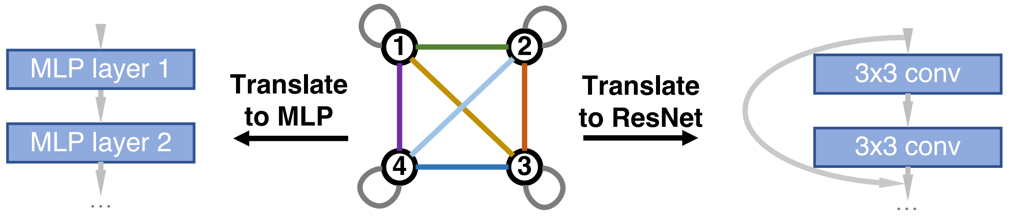
Specialty of GNNs:

(1) Graph structure is regarded as the **input instead of neural architecture**;

(2) Message functions are shared across all the edges to respect input graph's invariance properties.

Overview: **Methodology**

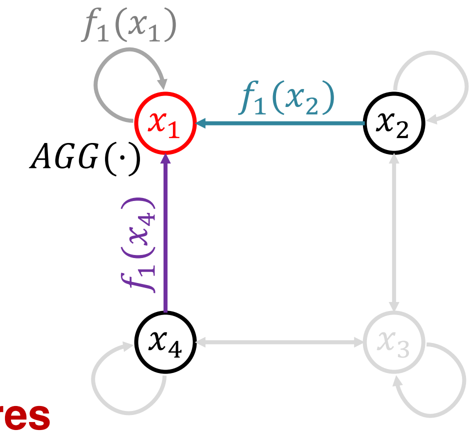
- A novel representation of neural networks: relational graphs
- Relational graphs can represent diverse neural architectures ←
  - Can represent architectures from MLP to ResNet



- Tools from network science → Graph structure vs NN performance

Relational Graphs → Diverse Architectures

Relational Graph



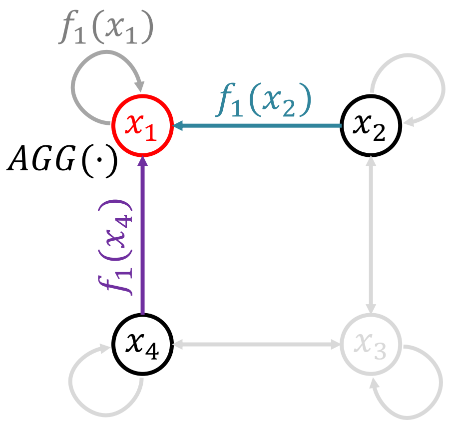
$$\mathbf{x}_v^{(r+1)} = \text{AGG}^{(r)}(\{f_v^{(r)}(\mathbf{x}_u^{(r)}), \forall u \in N(v)\})$$

The same relational graph → diverse architectures

4 Key components	**Fixed-width MLP**	**Variable-width MLP**	**ResNet-34**	**ResNet-34-sep**	**ResNet-50**
Node feature $\mathbf{x}_i$	Scalar: 1 dimension of data	Vector: multiple dimensions of data	Tensor: multiple channels of data	Tensor: multiple channels of data	Tensor: multiple channels of data
Message function $f_i(\cdot)$	Scalar multiplication	(Non-square) matrix multiplication	3×3 Conv	3×3 depth-wise and 1×1 Conv	3×3 and 1×1 Conv
Aggregation function $\text{AGG}(\cdot)$	$\sigma(\sum(\cdot))$	$\sigma(\sum(\cdot))$	$\sigma(\sum(\cdot))$	$\sigma(\sum(\cdot))$	$\sigma(\sum(\cdot))$
Number of rounds R	1 round per layer	1 round per layer	34 rounds with residual connections	34 rounds with residual connections	50 rounds with residual connections

Relational Graphs → Diverse Architectures

Relational Graph

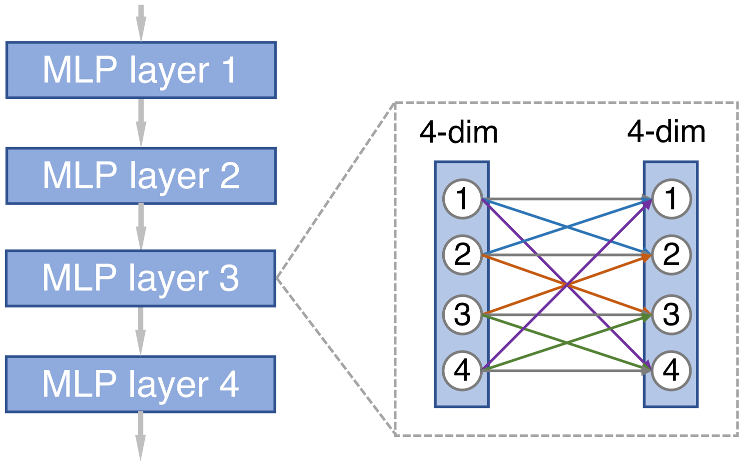


scalar

Node feature: $x_i \in \mathbb{R}$
Message: $f_i(x_j) = w_{ij}x_j$
Aggregation: $AGG(\cdot) = \sigma\sum(\cdot)$
Rounds: $R = 4$

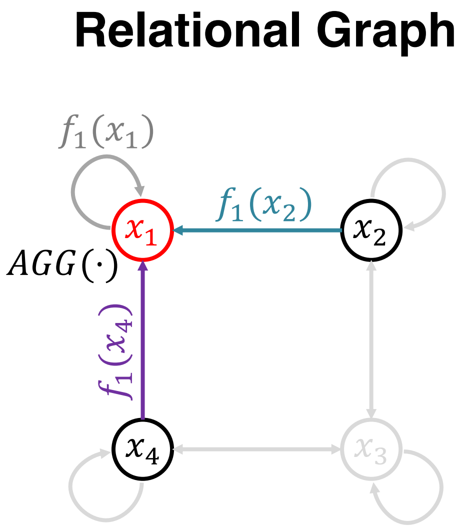
translate

4-layer 4-dim MLP



MLP as relational graph

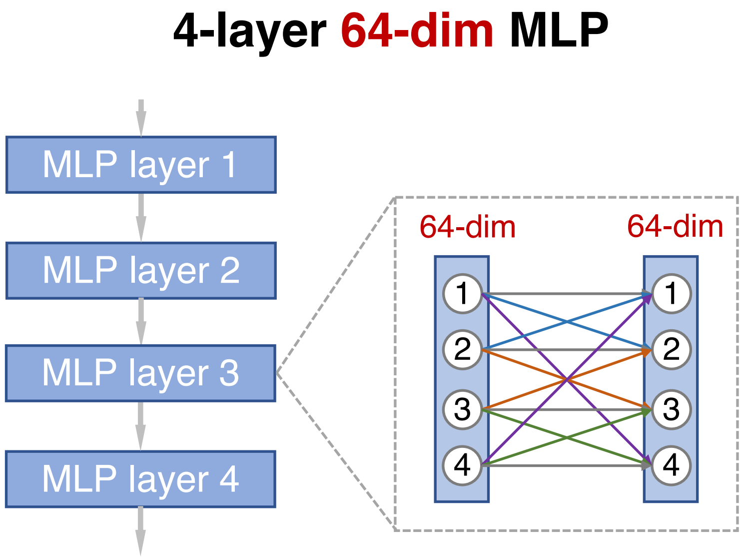
Relational Graphs → Diverse Architectures



vector

Node feature: $x_i \in \mathbb{R}^{16}$
Message: $f_i(x_j) = W_{ij}x_j$
Aggregation: $AGG(\cdot) = \sigma\sum(\cdot)$
Rounds: $R = 4$

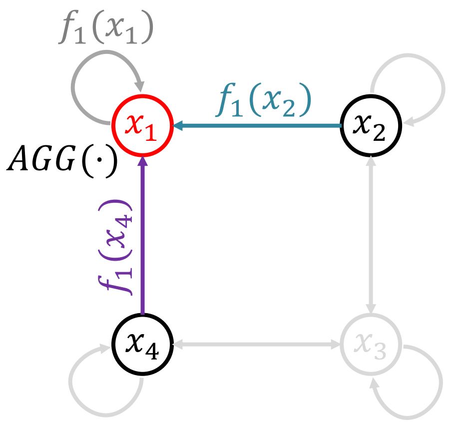
translate



MLP as relational graph

Relational Graphs → Diverse Architectures

Relational Graph

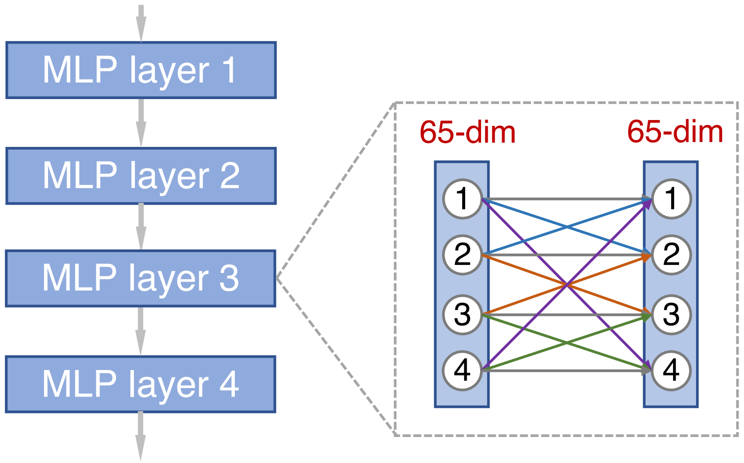


vector

Node feature: $x_{1,2,3} \in \mathbb{R}^{16}, x_4 \in \mathbb{R}^{17}$
Message: $f_i(x_j) = W_{ij}x_j$
Aggregation: $AGG(\cdot) = \sigma\sum(\cdot)$
Rounds: $R = 4$

translate

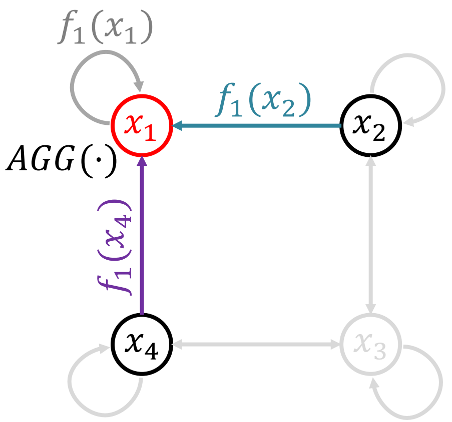
4-layer 65-dim MLP



MLP as relational graph

Relational Graphs → Diverse Architectures

Relational Graph

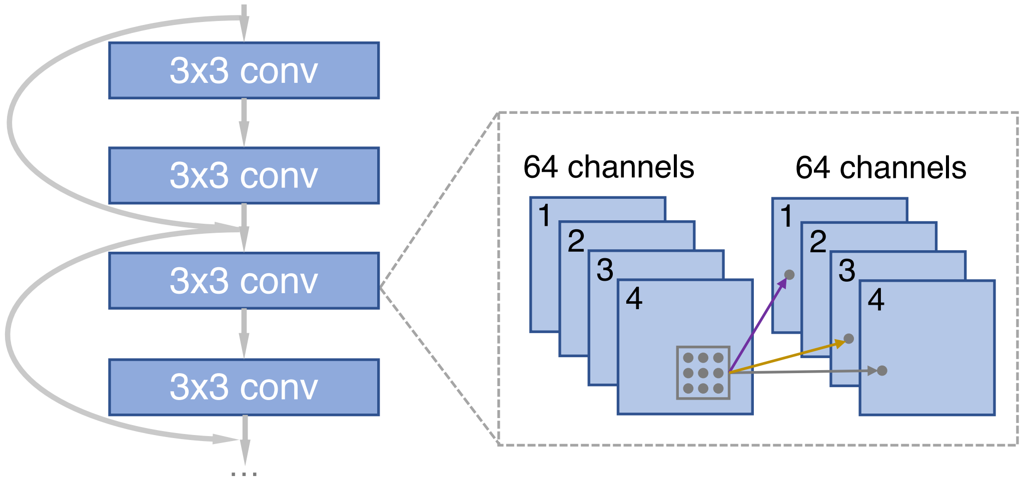


tensor

Node feature: $x_i \in \mathbb{R}^{16\times H\times W}$
Message: $f_i(x_j) = W_{ij} * x_j$
W_{ij} is 3x3 conv kernel
Aggregation: $AGG(\cdot) = \sigma\sum(\cdot)$
Rounds: $R = 34$

translate

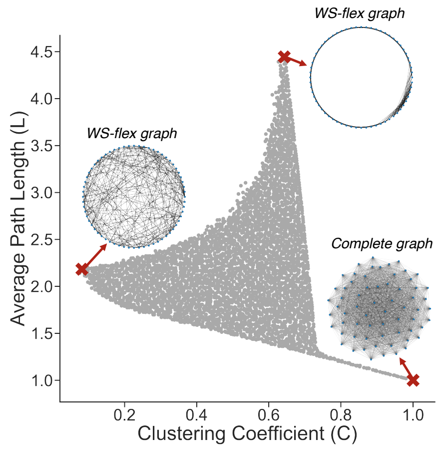
ResNet-34 64-channels



ResNet-34 as relational graph

Overview

- A novel representation of neural networks: relational graphs
- Relational graphs can represent diverse neural architectures
- Network science → Graph structure vs NN performance ←
  - Graph measures that characterize graph properties
  - Graph generators that generate diverse graphs
  - Control computational budget

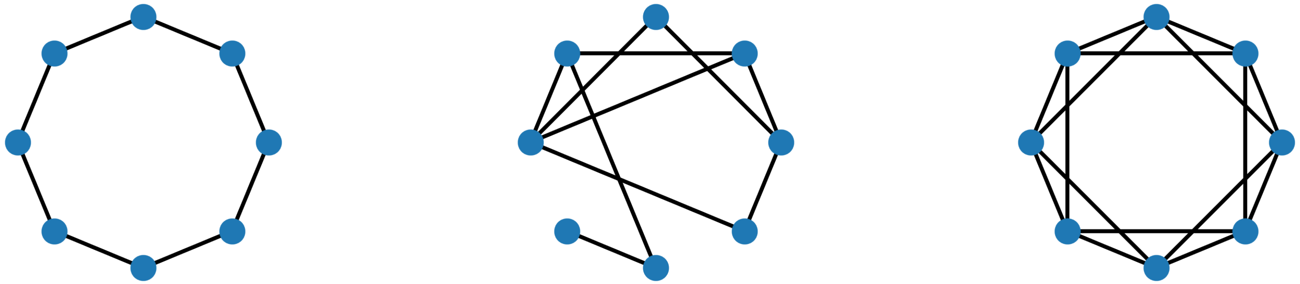


Measuring Graph Structure

- Graph measures:
  - Global: average path length (L)

    The average shortest path distance between any pair of nodes
  - Local: clustering coefficient (C)

    A measure of the degree to which nodes in a graph tend to cluster together

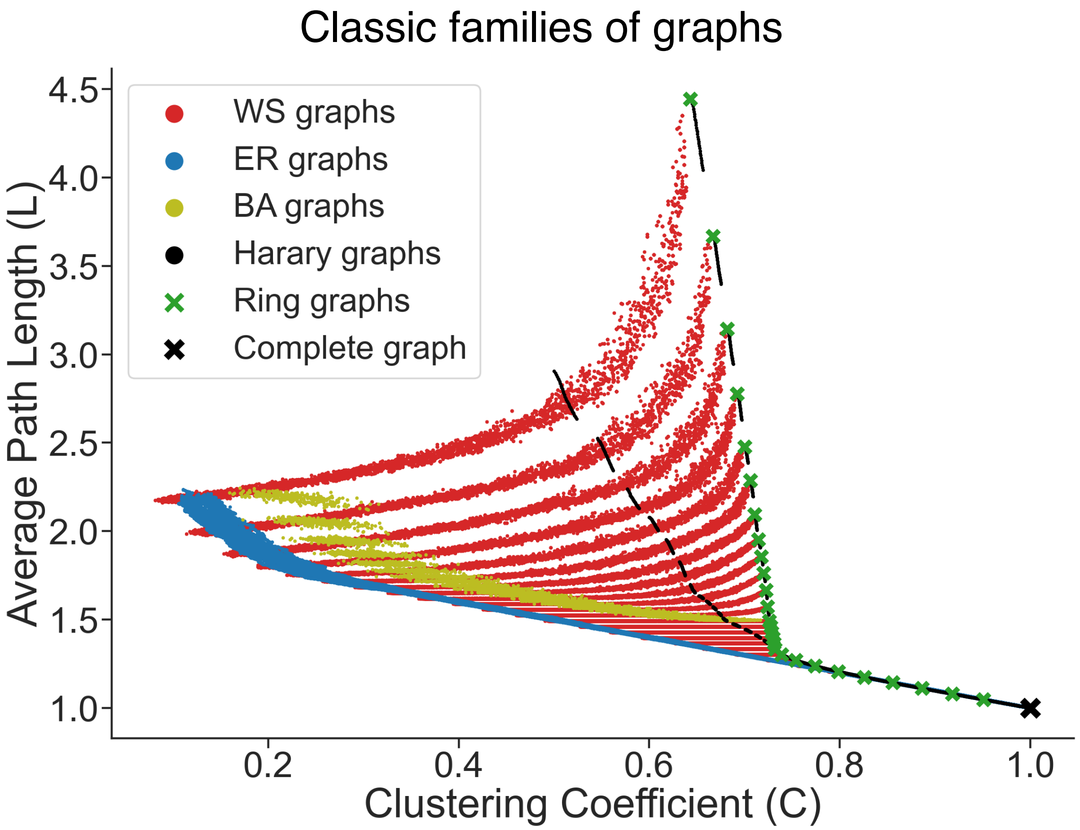


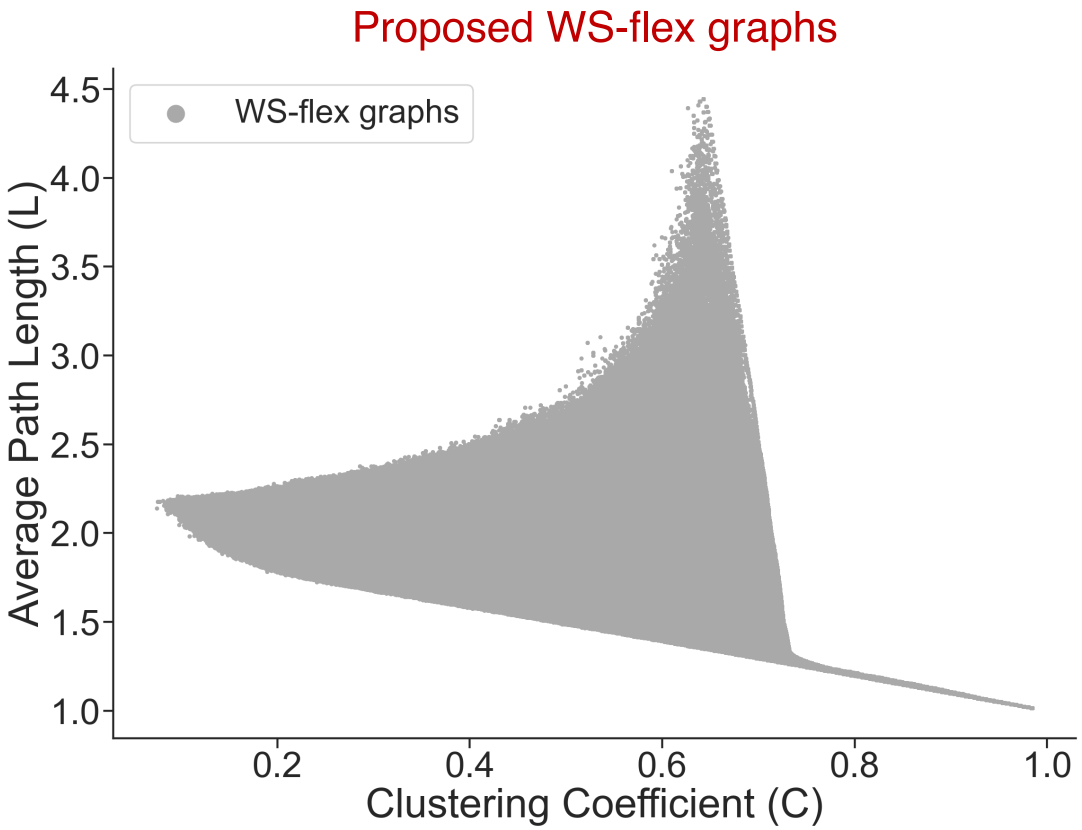
$L = 2.3$	$L = 2.0$	$L = 1.4$
$C = 0$	$C = 0.1$	$C = 0.5$

J. You, J. Leskovec, K. He, S. Xie, *Graph Structure of Neural Networks, ICML 2020*

Generating Diverse Graphs

- WS-flex graphs have a much better coverage

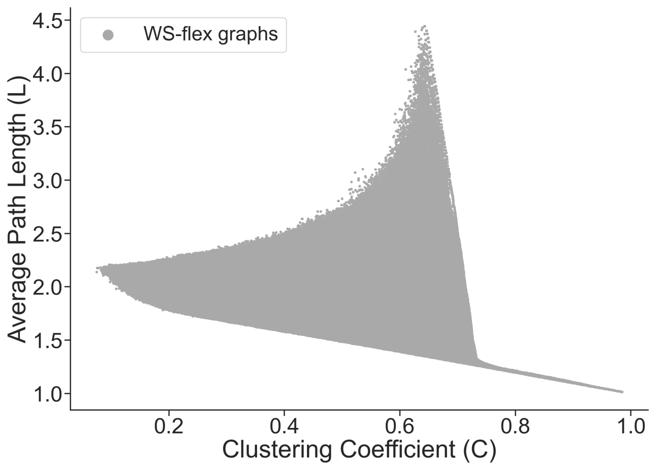


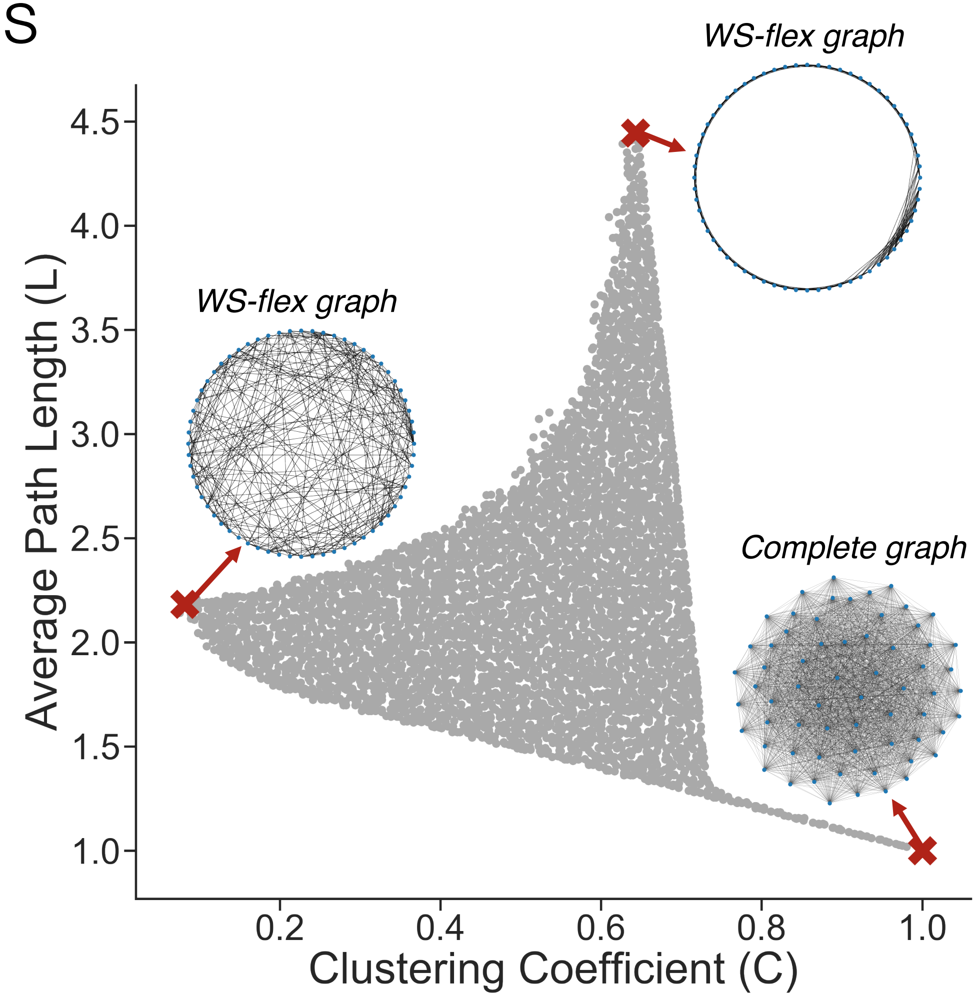


J. You, J. Leskovec, K. He, S. Xie, *Graph Structure of Neural Networks, ICML 2020*

Generating Diverse Graphs

- Visualization of WS-flex graphs





We Control Computational Budget

Relational graphs

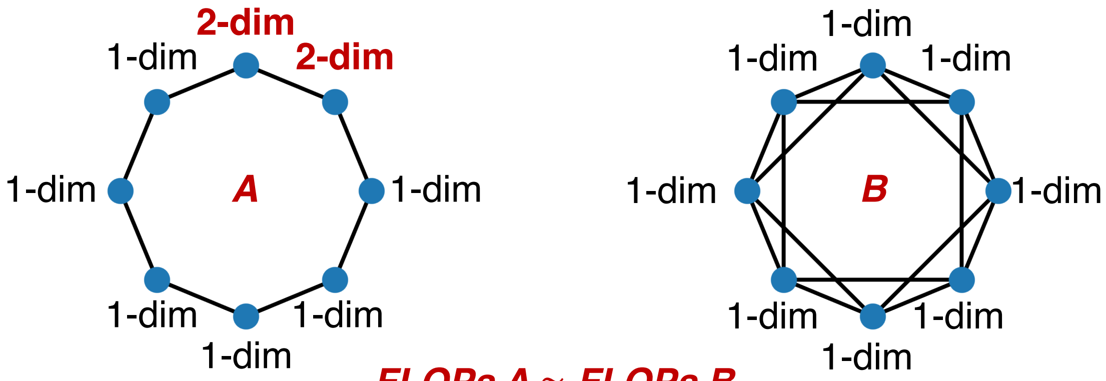
Total: 8-dim layer | **Total:** 8-dim layer

FLOPs A < FLOPs B

Have different computational FLOPs! ☹
FLOPs = O(num of edges)

Relational graphs

Total: 10-dim layer | **Total:** 8-dim layer



FLOPs A ≈ FLOPs B

Matched computational FLOPs! ☺
By controlling each node's dimension

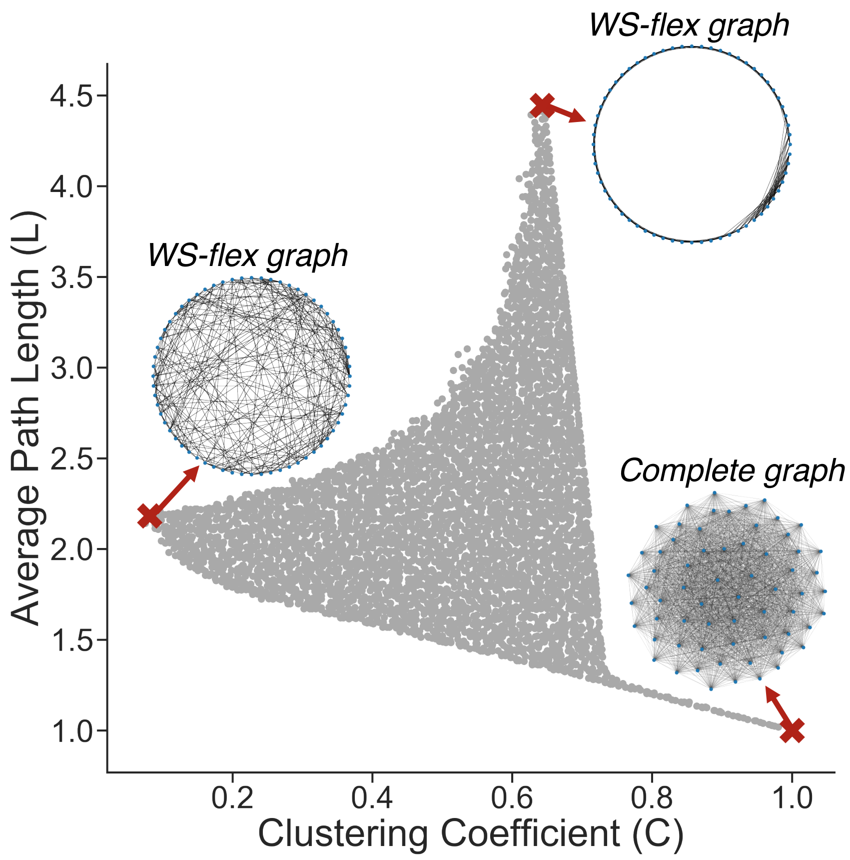
Experimental Setup

- 5-layer 512-dim MLPs on CIFAR-10
  - 3942 graphs, results averaged over 5 seeds
- CNNs & ResNet families & EfficientNet-B0 on ImageNet
  - 52 graphs per experiment, results averaged over 3 seeds
- Computational budgets in all experiments are controlled

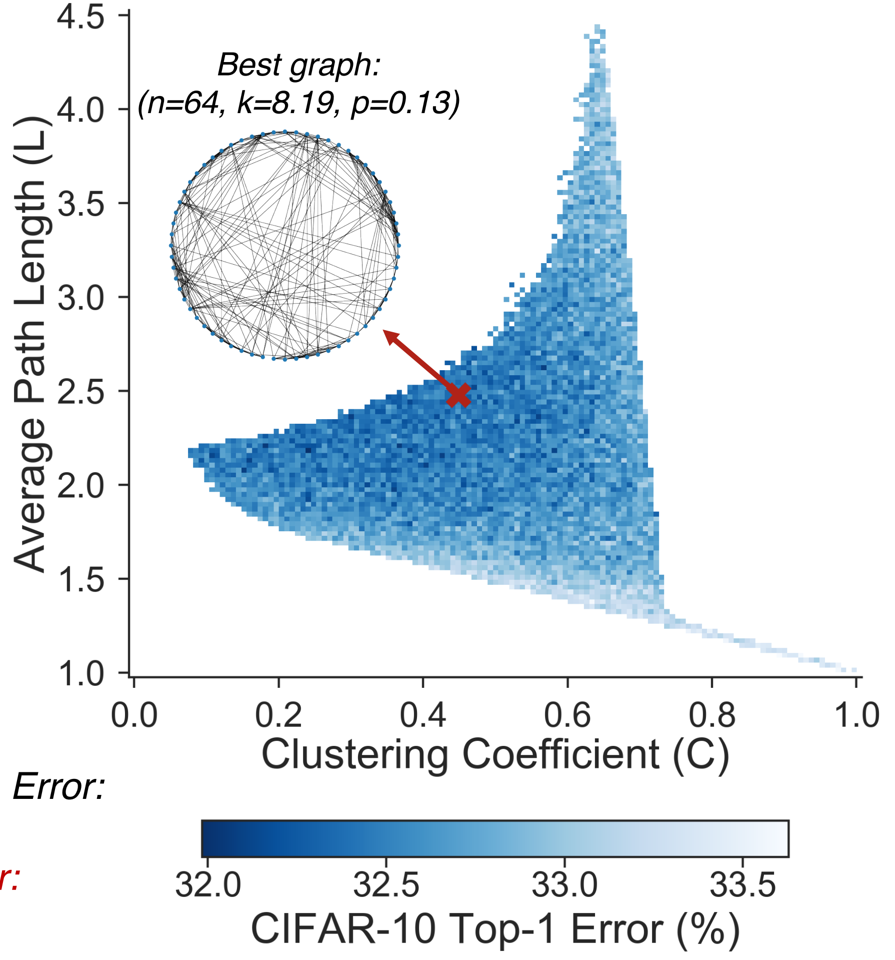
Overview: Findings

- Finding 1: Consistent Sweet Spot for top NNs across architectures
- Finding 2: NN Performance as a **smooth function** over graph measures
- Finding 3: Sweet spot can be **quickly identified**
- Finding 4: Top artificial NNs are **similar to real NNs**

Finding 1: **Sweet Spot** for Top NNs

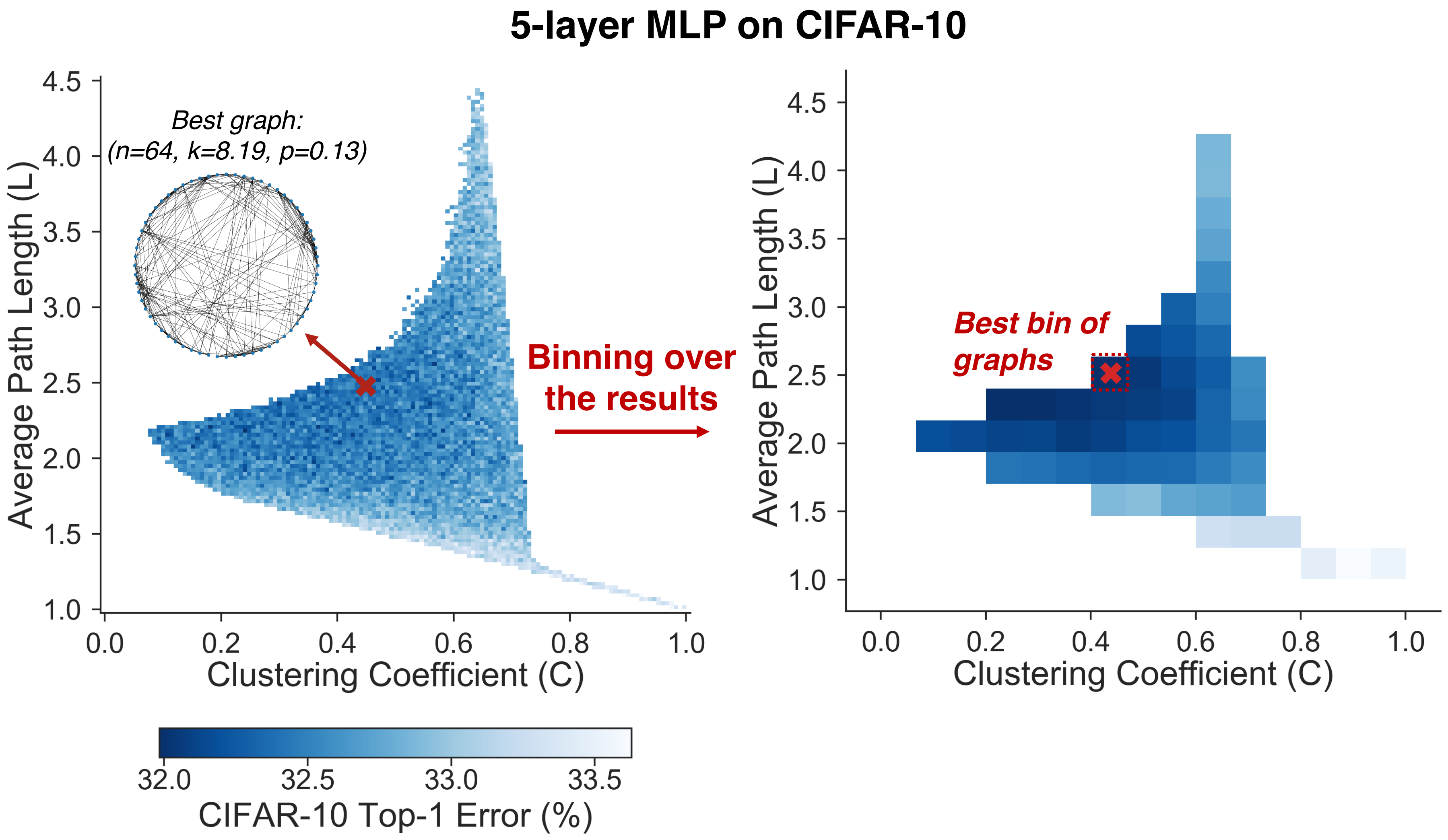


Translate to 5-layer MLP



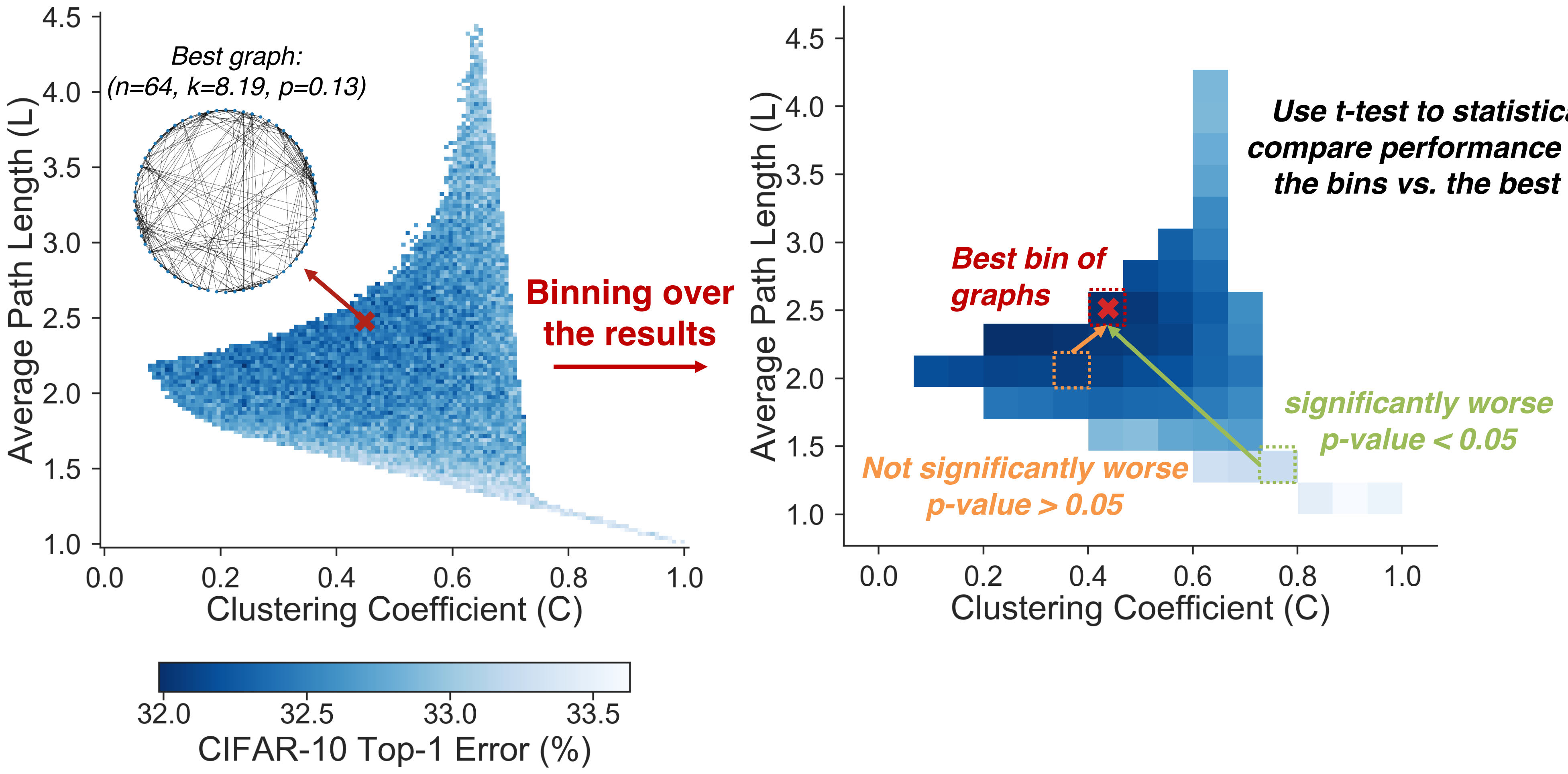
Complete graph Top-1 Error:
33.34 ± 0.36
Best graph Top-1 Error:
32.05 ± 0.14

Finding 1: Sweet Spot for Top NNs



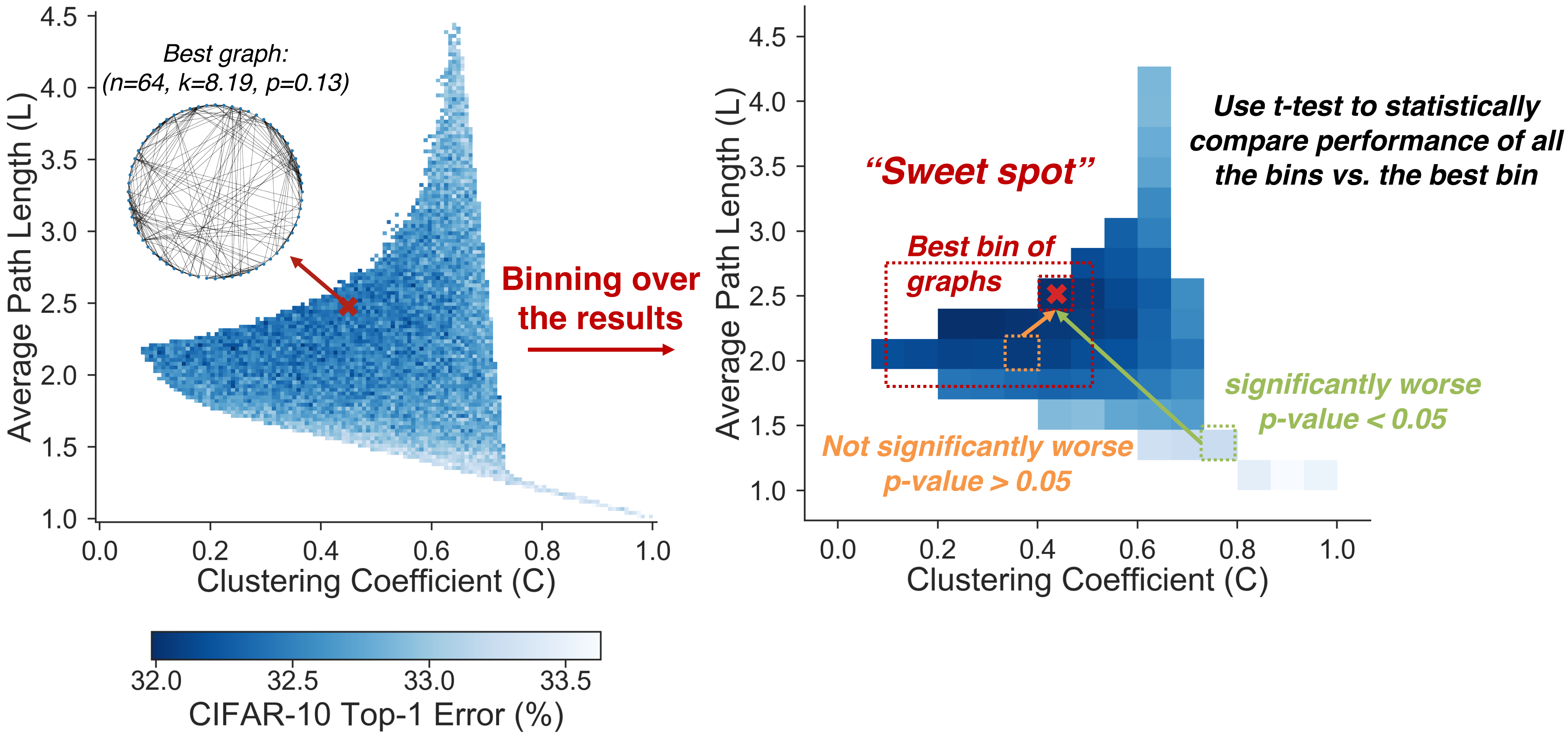
Finding 1: Sweet Spot for Top NNs

5-layer MLP on CIFAR-10



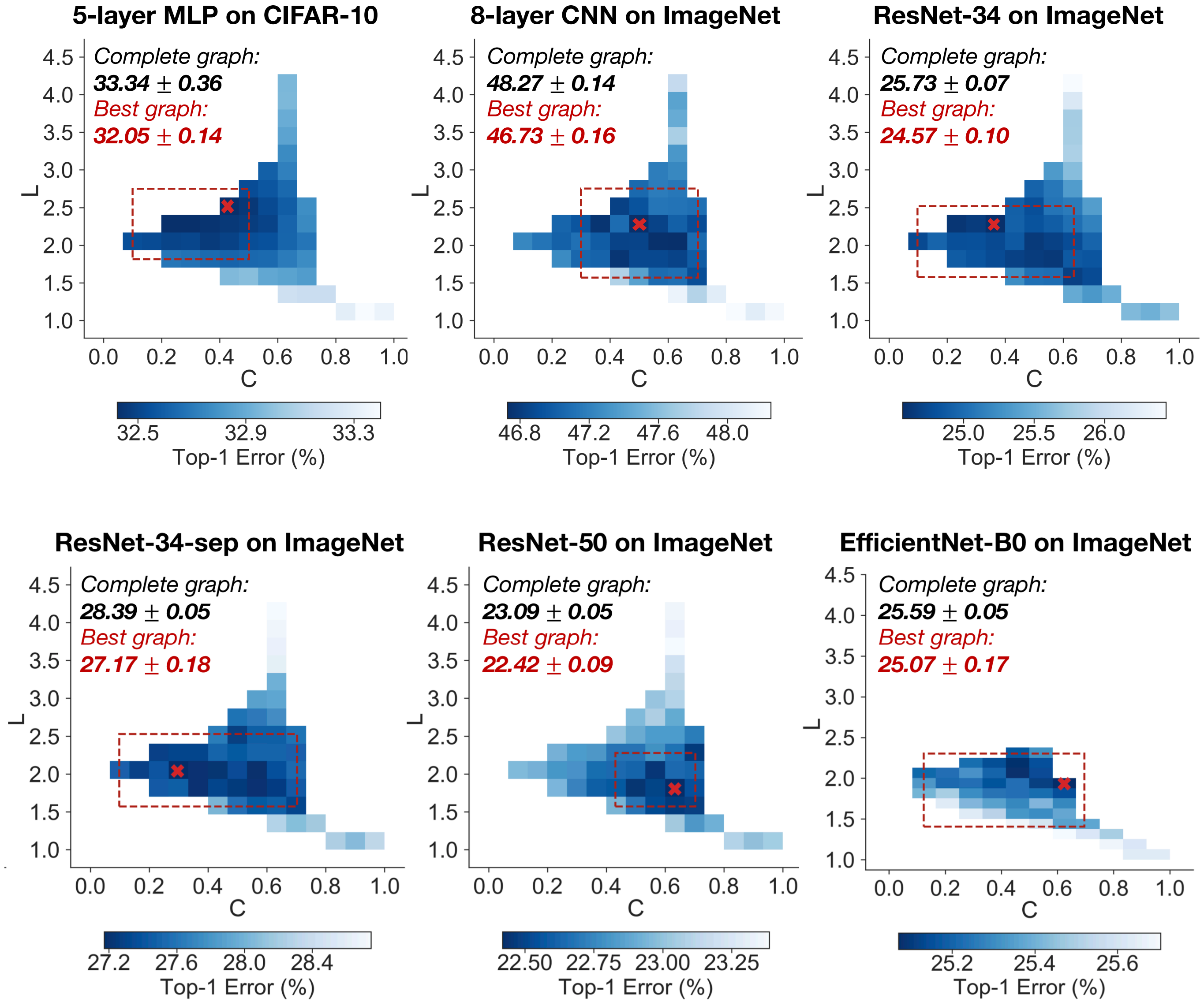
Finding 1: Sweet Spot for Top NNs

5-layer MLP on CIFAR-10



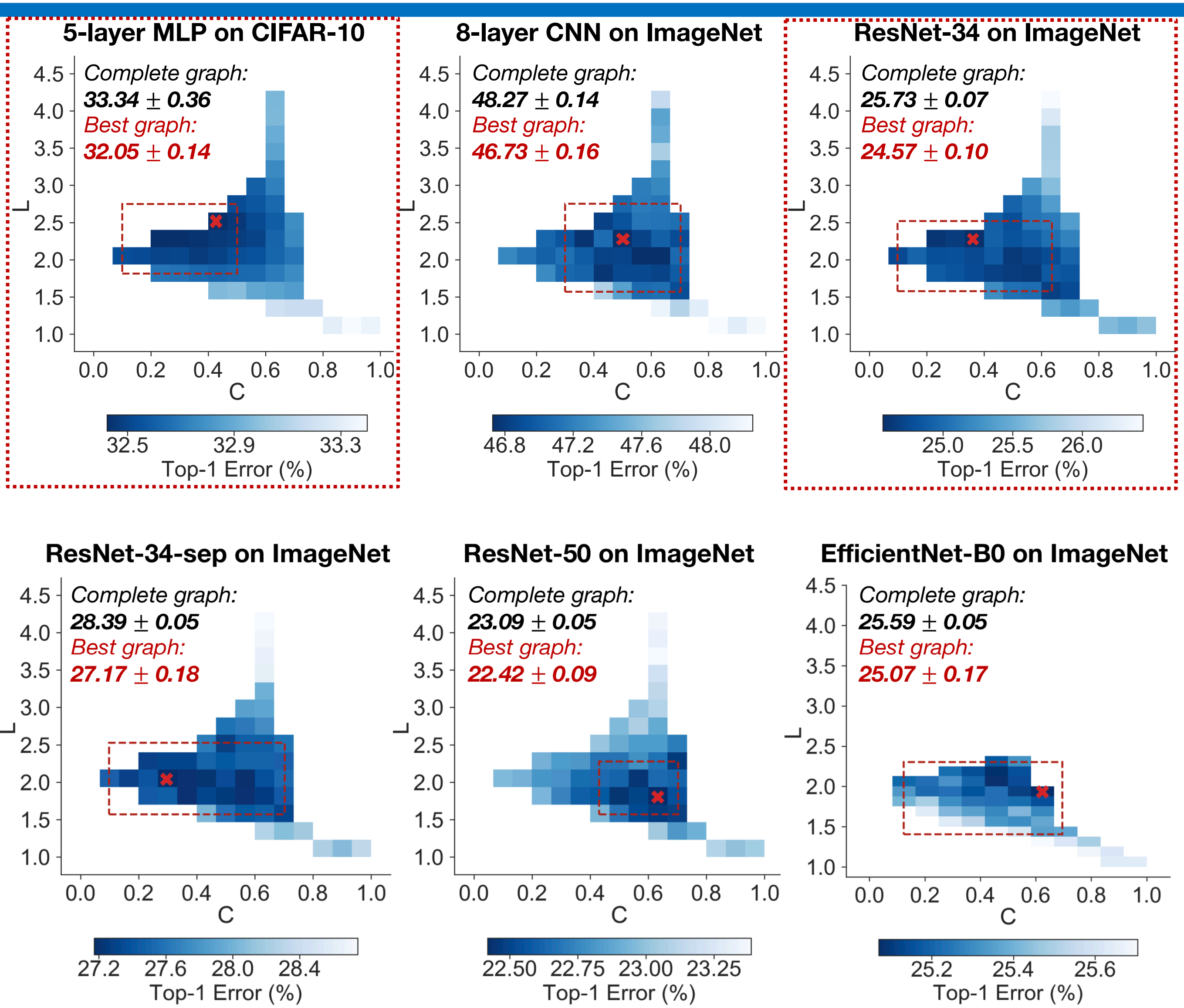
Finding 1: Consistent Sweet Spot for Top NNs

A consistent sweet spot
$C \in [0.43, 0.50]$
$L \in [1.82, 2.28]$



Finding 1: Consistent Sweet Spot for Top NNs

2.89e6 FLOPS



3.66e9 FLOPS

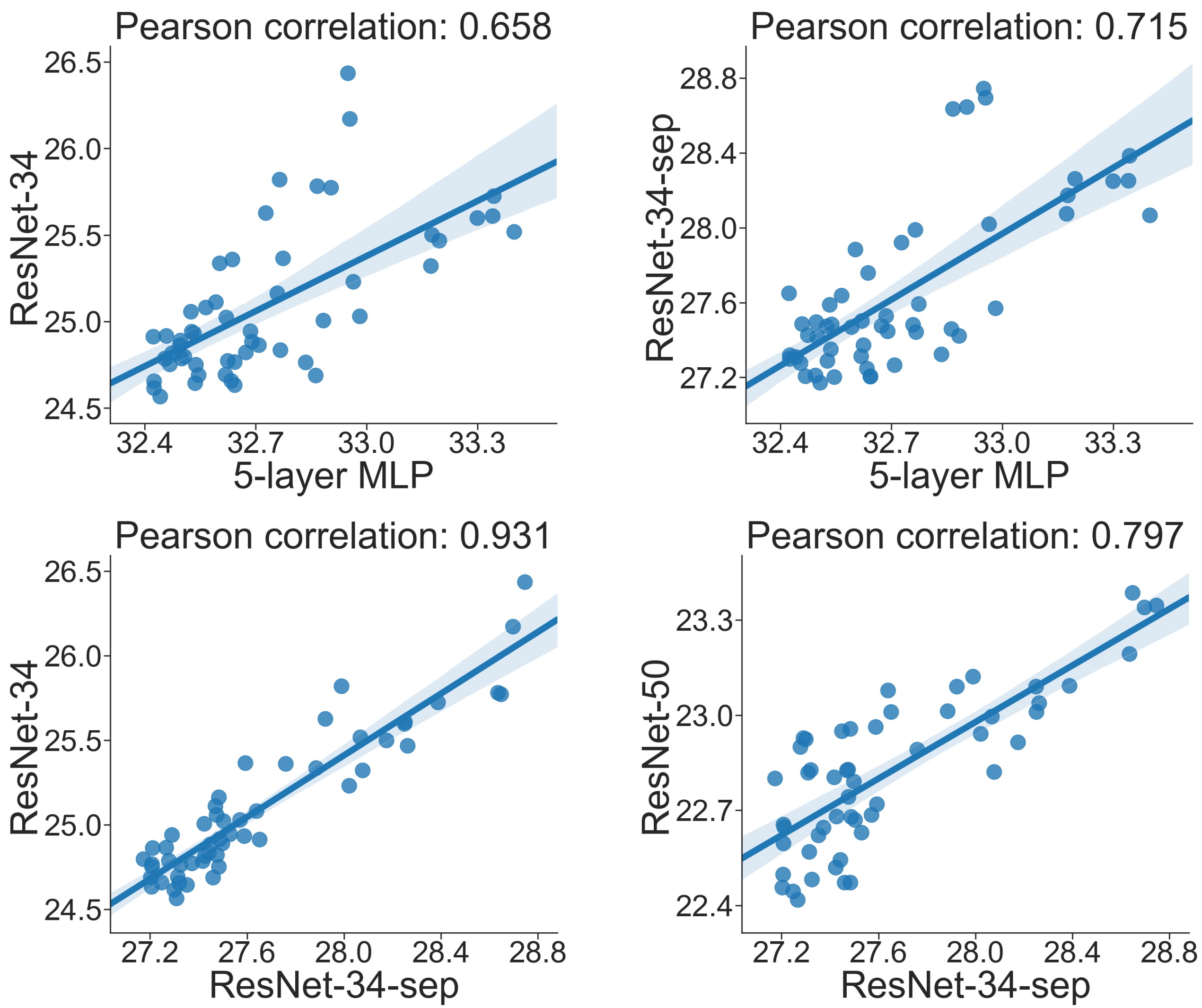
~1000x FLOPS difference
But similar sweet spot!

A consistent sweet spot

$C \in [0.43, 0.50]$

$L \in [1.82, 2.28]$

Finding 1: **Consistent Sweet Spot** for Top NNs

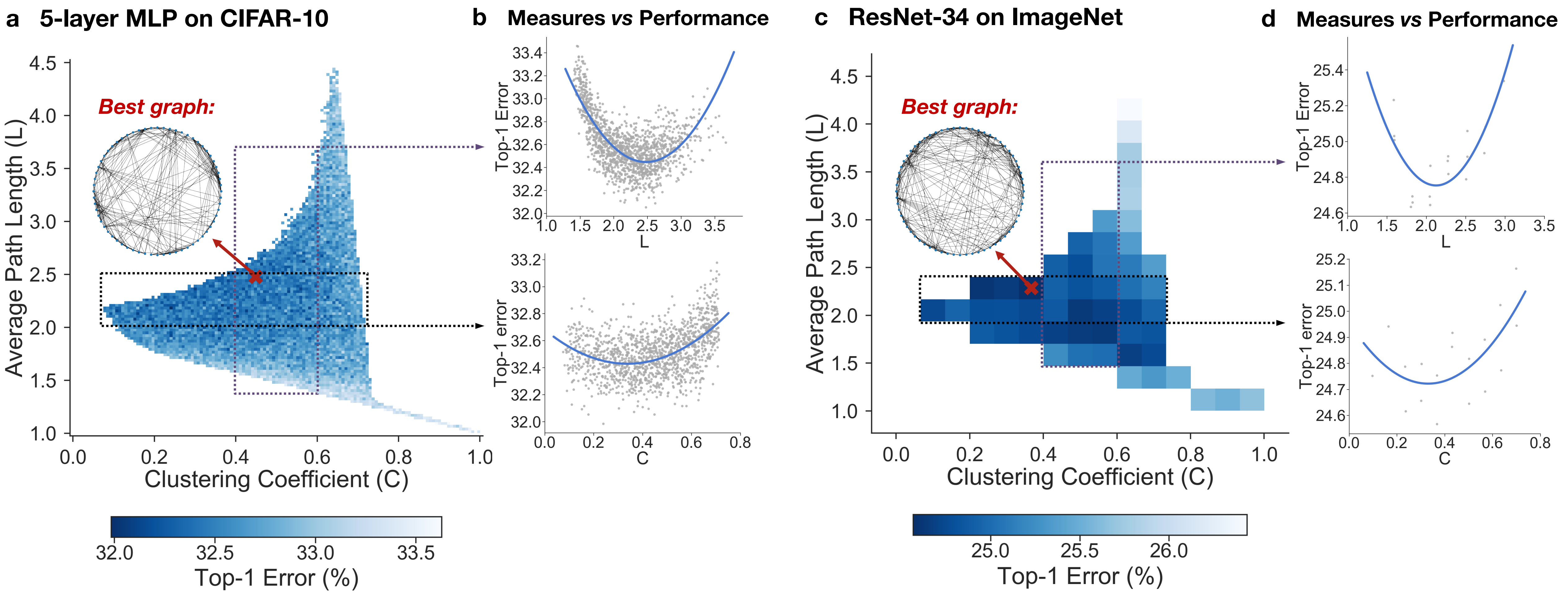


Quantitative consistency across architectures

Overview: **Findings**

- **Finding 1: Consistent Sweet Spot** for top NNs across architectures
- **Finding 2:** NN Performance as a **smooth function** over graph measures
- **Finding 3:** Sweet spot can be **quickly identified**
- **Finding 4:** Top artificial NNs are **similar to real NNs**

Finding 2: NN performance as a **smooth function** over graph measures



Overview: **Findings**

- Finding 1: **Consistent Sweet Spot** for top NNs across architectures

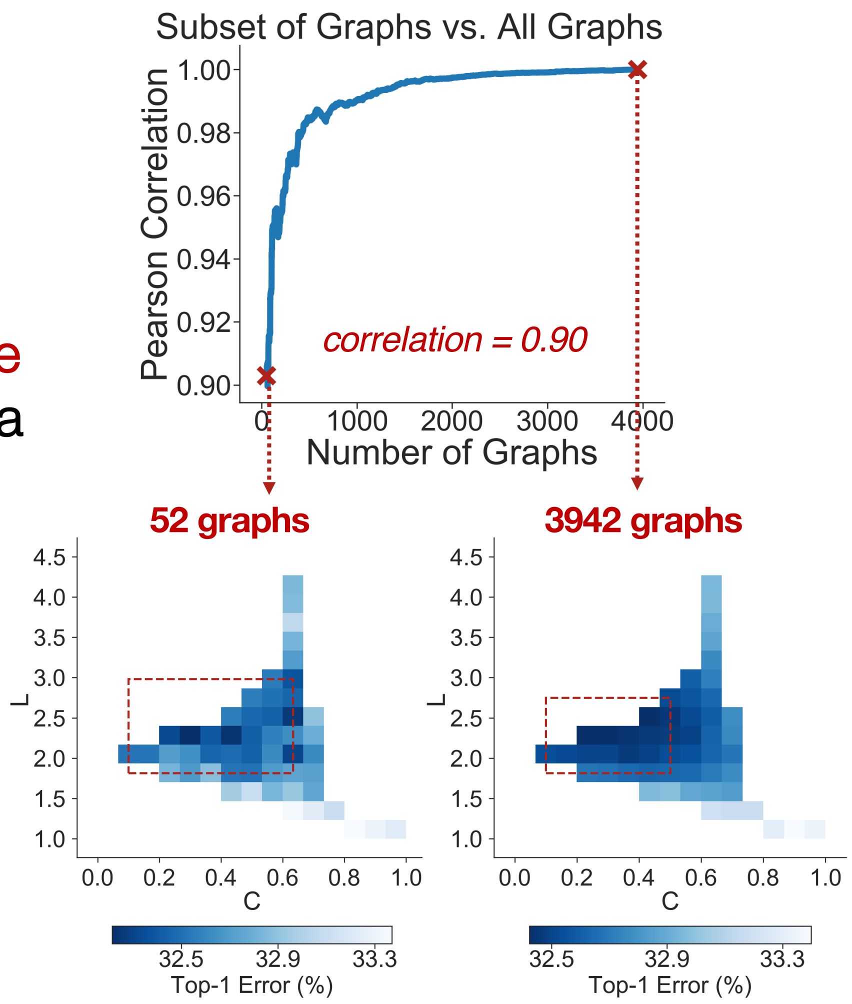
- Finding 2: NN Performance as a **smooth function** over graph measures

- Finding 3: Sweet spot can be **quickly identified**

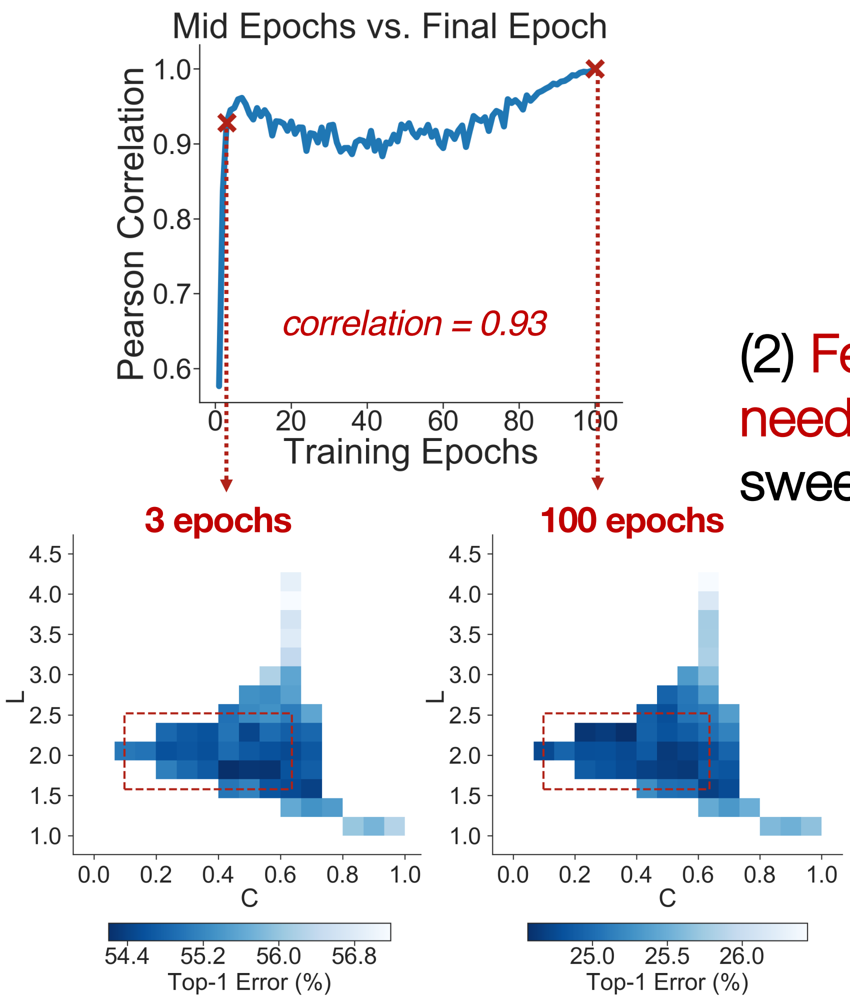
- Finding 4: Top artificial NNs are **similar to real NNs**

Finding 3: Sweet spot can be **quickly identified**

(1) Few graphs are needed to locate a sweet spot



5-layer MLP on CIFAR-10



ResNet-34 on ImageNet

(2) Few epochs are needed to locate a sweet spot

Overview: **Findings**

- **Finding 1: Consistent Sweet Spot** for top NNs across architectures

- **Finding 2:** NN Performance as a **smooth function** over graph measures

- **Finding 3:** Sweet spot can be **quickly identified**

- **Finding 4:** Top artificial NNs are similar to real NNs

Finding 4: Top artificial NNs are **similar to real NNs**

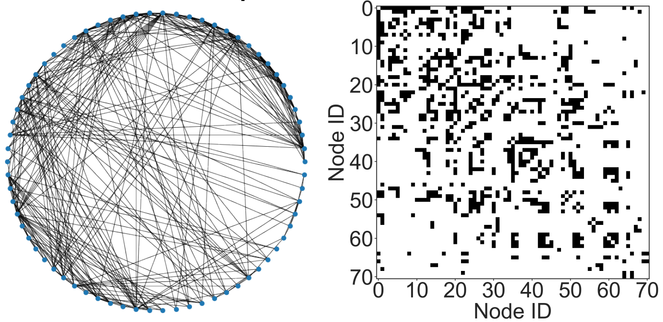
a

Graph	Path (L)	Clustering (C)	CIFAR-10 Error (%)
Complete graph	1.00	1.00	33.34 ± 0.36
Cat cortex	1.81	0.55	33.01 ± 0.22
Macaque visual cortex	1.73	0.53	32.78 ± 0.21
Macaque whole cortex	**2.38**	**0.46**	32.77 ± 0.14
Consistent sweet spot across neural architectures	**1.82-2.28**	**0.43-0.50**	32.50 ± 0.33
Best 5-layer MLP	**2.48**	**0.45**	32.05 ± 0.14

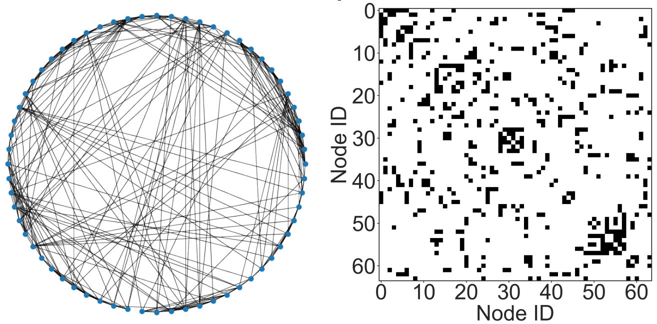
(1) Best graphs we found are similar to biological neural networks

b

Biological neural network:
Macaque whole cortex



Artificial neural network:
Best 5-layer MLP



Finding 4: Top artificial NNs are **similar to real NNs**

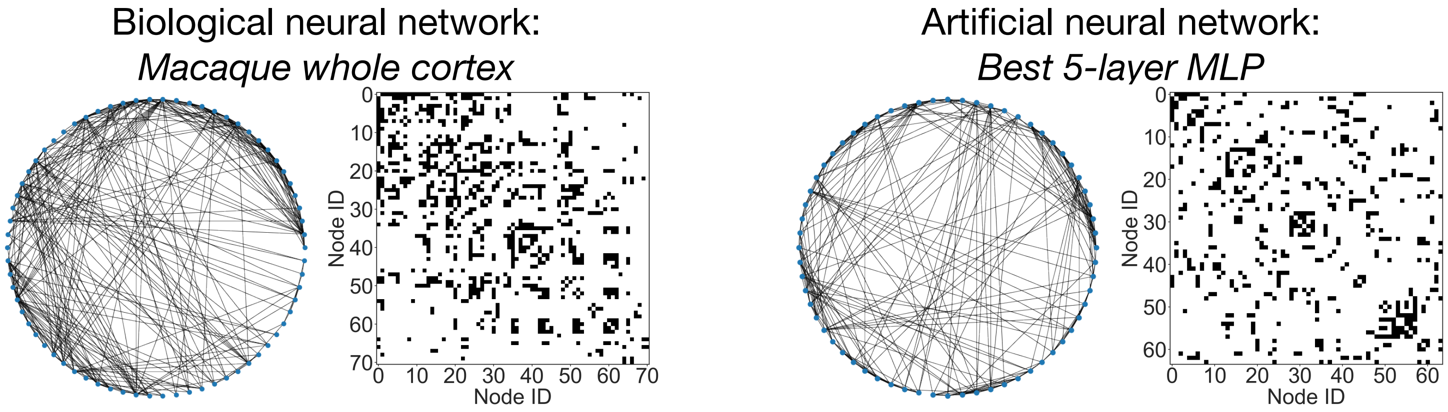
a

Graph	Path (L)	Clustering (C)	CIFAR-10 Error (%)
Complete graph	1.00	1.00	33.34 ± 0.36
Cat cortex	1.81	0.55	33.01 ± 0.22
Macaque visual cortex	1.73	0.53	32.78 ± 0.21
Macaque whole cortex	**2.38**	**0.46**	32.77 ± 0.14
Consistent sweet spot across neural architectures	**1.82-2.28**	**0.43-0.50**	32.50 ± 0.33
Best 5-layer MLP	**2.48**	**0.45**	32.05 ± 0.14

(2) Translate biological networks to MLP yields good performance

(1) Best graphs we found are similar to biological neural networks

b



Finding 4: Top artificial NNs are **similar to real NNs**

a

Graph	**Path (L)**	**Clustering (C)**	**CIFAR-10 Error (%)**
Complete graph	1.00	1.00	33.34 ± 0.36
Cat cortex	1.81	0.55	33.01 ± 0.22
Macaque visual cortex	1.73	0.53	32.78 ± 0.21
Macaque whole cortex	**2.38**	**0.46**	32.77 ± 0.14
Consistent sweet spot across neural architectures	**1.82-2.28**	**0.43-0.50**	32.50 ± 0.33
Best 5-layer MLP	**2.48**	**0.45**	32.05 ± 0.14

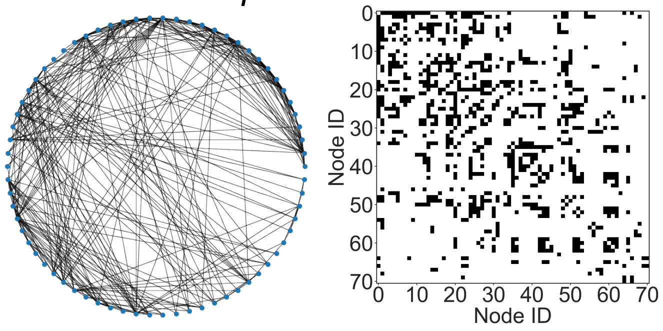
(2) Translate biological networks to MLP yields good performance

(1) Best graphs we found are similar to biological neural networks

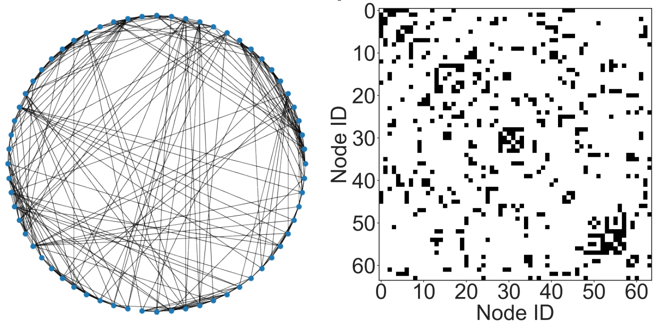
More advanced bio networks → better performed deep networks?

b

Biological neural network:
Macaque whole cortex

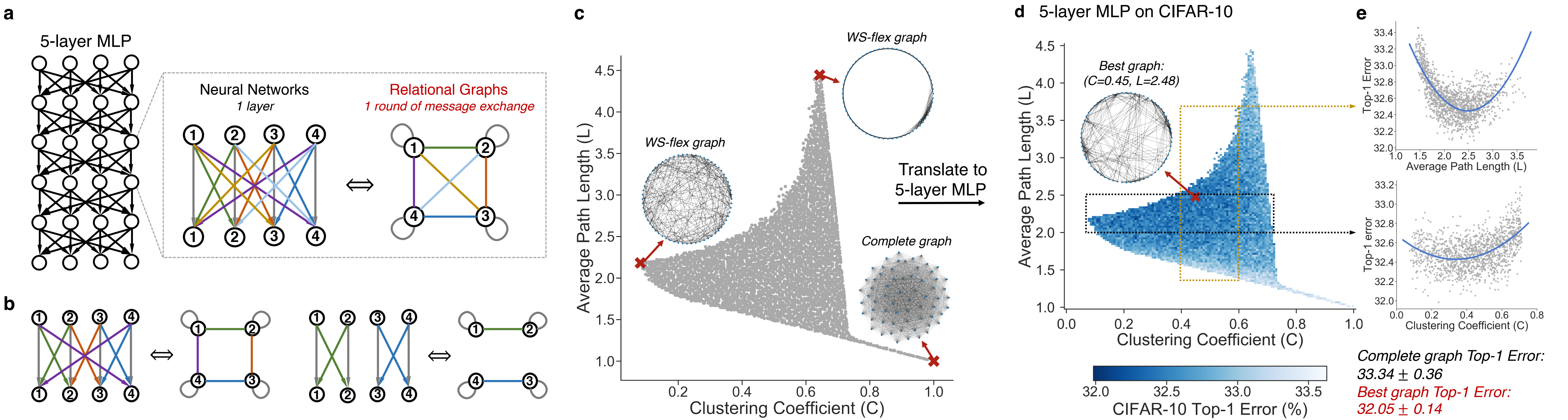


Artificial neural network:
Best 5-layer MLP



Conclusions

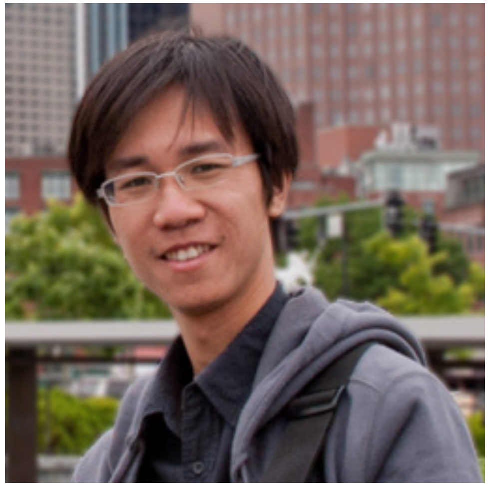
- A new transition from studying conventional computation architecture to studying ***graph structure* of neural networks**.
- Well-established methodologies from **network science and neuroscience** could contribute to **understanding and designing deep neural networks**.



Graph Structure of Neural Networks

Jiaxuan You

Jure Leskovec



Kaiming He

Saining Xie

FACEBOOK