Graph Structure of Neural Networks

Jiaxuan You¹, Jure Leskovec¹, Kaiming He², Saining Xie²

Stanford University¹
Facebook AI Research²
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?

...
Overview: Methodology

- A novel representation of neural networks: relational graphs

- Relational graphs can represent diverse neural architectures

- Tools from network science $\rightarrow$ Graph structure vs NN performance
Overview: Key Findings

- **Consistent “Sweet Spot”** for top NNs across architectures

- **Top artificial NNs are similar to real biological NNs**

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

- 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 \(\rightarrow\) Graph structure vs NN performance
Background: NNs as Computational Graphs

MLP model

MLP layer

MLP layer

MLP layer

MLP layer

Computational graphs

Computational graphs

1 layer

1 2 3 4

1 2 3 4
Background: Related Work

Existing graph-based architecture design approaches focus on computational graphs

Deep Expander Networks, Prabhu et al., 2018
RandWire, Xie et al., 2019
NAS-Bench-101, Ying et al., 2019

Generate **bipartite graphs** over neurons

Generate **directed acyclic graphs** over NN layers

J. You, J. Leskovec, K. He, S. Xie, Graph Structure of Neural Networks, ICML 2020
Limitations: NNs as Computational Graphs

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

Computational graphs
1 layer

Relational Graphs
1 round of message exchange
Our Approach: Relational Graphs

Computational graphs
1 layer

Relational Graphs
1 round of message exchange
Benefits of Relational Graphs

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

Computational graphs
- 1 layer

Relational Graphs
- 1 round of message exchange

Biological neural network: Macaque whole cortex

Artificial neural network: Best 5-layer MLP

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

J. You, J. Leskovec, K. He, S. Xie, Graph Structure of Neural Networks, ICML 2020
Neural Computation as \textbf{Message Exchange}

\textbf{Computational graphs}
\textit{Directed message flow}

\begin{center}
\begin{tikzpicture}
\node[circle,draw,fill=white] (1) at (0,0) {1};
\node[circle,draw,fill=white] (2) at (1,1) {2};
\node[circle,draw,fill=white] (3) at (2,0) {3};
\node[circle,draw,fill=white] (4) at (1,-1) {4};
\node[circle,draw,fill=white] (A) at (2,2) {1};
\node[circle,draw,fill=white] (B) at (3,1) {2};
\node[circle,draw,fill=white] (C) at (4,0) {3};
\node[circle,draw,fill=white] (D) at (3,-1) {4};
\draw[->,thick] (1) -- (2);
\draw[->,thick] (1) -- (3);
\draw[->,thick] (1) -- (4);
\draw[->,thick] (2) -- (1);
\draw[->,thick] (2) -- (3);
\draw[->,thick] (2) -- (4);
\draw[->,thick] (3) -- (1);
\draw[->,thick] (3) -- (2);
\draw[->,thick] (3) -- (4);
\draw[->,thick] (4) -- (1);
\draw[->,thick] (4) -- (2);
\draw[->,thick] (4) -- (3);
\node at (0.5,0.5) {$AGG(\cdot)$};
\node at (1.5,1.5) {$AGG(\cdot)$};
\node at (2.5,0.5) {$AGG(\cdot)$};
\node at (1.5,-0.5) {$AGG(\cdot)$};
\end{tikzpicture}
\end{center}

\textbf{Relational graphs}
\textit{Bi-Directed message exchange}

\begin{center}
\begin{tikzpicture}
\node[circle,draw,fill=white] (1) at (0,0) {1};
\node[circle,draw,fill=white] (2) at (1,1) {2};
\node[circle,draw,fill=white] (3) at (2,0) {3};
\node[circle,draw,fill=white] (4) at (1,-1) {4};
\node[circle,draw,fill=white] (A) at (2,2) {1};
\node[circle,draw,fill=white] (B) at (3,1) {2};
\node[circle,draw,fill=white] (C) at (4,0) {3};
\node[circle,draw,fill=white] (D) at (3,-1) {4};
\draw[->,thick,bend left] (1) to (2);
\draw[->,thick,bend left] (1) to (4);
\draw[->,thick,bend left] (2) to (1);
\draw[->,thick,bend left] (2) to (3);
\draw[->,thick,bend left] (3) to (2);
\draw[->,thick,bend left] (3) to (4);
\draw[->,thick,bend left] (4) to (3);
\node at (0.5,0.5) {$AGG(\cdot)$};
\node at (1.5,1.5) {$AGG(\cdot)$};
\node at (2.5,0.5) {$AGG(\cdot)$};
\node at (1.5,-0.5) {$AGG(\cdot)$};
\end{tikzpicture}
\end{center}
A 5-layer Neural network

A relational graph with 5 rounds of message exchange

This is how Graph Neural Networks compute embeddings!
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

The same relational graph → diverse architectures

<table>
<thead>
<tr>
<th>Key components</th>
<th>Fixed-width MLP</th>
<th>Variable-width MLP</th>
<th>ResNet-34</th>
<th>ResNet-34-sep</th>
<th>ResNet-50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Node feature $x_i$</td>
<td>Scalar: 1 dim. of data</td>
<td>Vector: multiple dim. of data</td>
<td>Tensor: multiple channels of data</td>
<td>Tensor: multiple channels of data</td>
<td>Tensor: multiple channels of data</td>
</tr>
<tr>
<td>Message function $f_i(\cdot)$</td>
<td>Scalar multiplication</td>
<td>(Non-square) matrix multiplication</td>
<td>3×3 Conv</td>
<td>3×3 depth-wise and 1×1 Conv</td>
<td>3×3 and 1×1 Conv</td>
</tr>
<tr>
<td>Aggregation function $AGG(\cdot)$</td>
<td>$\sigma(\sum(\cdot))$</td>
<td>$\sigma(\sum(\cdot))$</td>
<td>$\sigma(\sum(\cdot))$</td>
<td>$\sigma(\sum(\cdot))$</td>
<td>$\sigma(\sum(\cdot))$</td>
</tr>
<tr>
<td>Number of rounds $R$</td>
<td>1 round per layer</td>
<td>1 round per layer</td>
<td>34 rounds with residual connections</td>
<td>34 rounds with residual connections</td>
<td>50 rounds with residual connections</td>
</tr>
</tbody>
</table>
Relational Graphs ➔ Diverse Architectures

Relational Graph

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

MLP as relational graph

4-layer 4-dim MLP

MLP layer 1
MLP layer 2
MLP layer 3
MLP layer 4

scalar

Translate
Relational Graphs ➔ Diverse Architectures

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

MLP as relational graph
Relational Graphs → Diverse Architectures

**Relational Graph**

- Node feature: \( x_1, x_2, x_3, x_4 \in \mathbb{R}^{16}, x_4 \in \mathbb{R}^{17} \)
- Message: \( f_i(x_j) = W_{ij}x_j \)
- Aggregation: \( AGG(\cdot) = \sigma \sum(\cdot) \)
- Rounds: \( R = 4 \)

**MLP as relational graph**

4-layer 65-dim MLP

- MLP layer 1
- MLP layer 2
- MLP layer 3
- MLP layer 4

MLP layer 2

\[ \begin{align*}
\text{vector} \quad &\downarrow \\
\text{translate} \quad &\rightarrow \\
\text{MLP layer 1} \quad &\downarrow \\
\text{MLP layer 2} \quad &\downarrow \\
\text{MLP layer 3} \quad &\downarrow \\
\text{MLP layer 4} \quad &\downarrow \\
\end{align*} \]
Relational Graphs ➔ Diverse Architectures

Relational Graph

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

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 $\rightarrow$ Graph structure vs NN performance $\leftarrow$

- 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

\[
L = 2.3 \\
C = 0
\]

\[
L = 2.0 \\
C = 0.1
\]

\[
L = 1.4 \\
C = 0.5
\]
Generating Diverse Graphs

- WS-flex graphs have a much better coverage
Generating Diverse Graphs

- Visualization of WS-flex graphs
We Control Computational Budget

Relational graphs

**Total: 8-dim layer**

- **A**
  - 1-dim
  - 1-dim
  - 1-dim
  - 1-dim
  - 1-dim

- **B**
  - 1-dim
  - 1-dim
  - 1-dim
  - 1-dim

FLOPs $A < $FLOPs $B$

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

Relational graphs

**Total: 10-dim layer**

- **A**
  - 2-dim
  - 1-dim
  - 1-dim
  - 1-dim
  - 1-dim

- **B**
  - 1-dim
  - 1-dim
  - 1-dim
  - 1-dim

FLOPs $A \approx $FLOPs $B$

Matched computational FLOPs! 😊
By controlling each node’s dimension

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

WS-flex graph

Complete graph

Best graph:
(n=64, k=8.19, p=0.13)
Finding 1: **Sweet Spot for Top NNs**

Best graph:
(n=64, k=8.19, p=0.13)

Binning over the results

5-layer MLP on CIFAR-10
Finding 1: Sweet Spot for Top NNs

Best graph: 
\( (n=64, k=8.19, p=0.13) \)

Not significantly worse 
\( p\)-value > 0.05

Significantly worse 
\( p\)-value < 0.05

Use t-test to statistically compare performance of all the bins vs. the best bin
Finding 1: Sweet Spot for Top NNs

Best graph: $(n=64, k=8.19, p=0.13)$

Not significantly worse
$p$-value > 0.05

significantly worse
$p$-value < 0.05

Use t-test to statistically compare performance of all the bins vs. the best bin

Binning over the results

“Sweet spot”

Best bin of graphs

Not significantly worse
$p$-value > 0.05

significantly worse
$p$-value < 0.05

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

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

Best graph:

- 5-layer MLP on CIFAR-10
- Measures vs Performance
- ResNet-34 on ImageNet
- Measures vs Performance
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

(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

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

(1) Best graphs we found are similar to biological neural networks
Finding 4: Top artificial NNs are similar to real NNs

### Table

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

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

(2) Translate biological networks to MLP yields good performance
Finding 4: Top artificial NNs are similar to real NNs

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

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

(2) Translate biological networks to MLP yields good performance

More advanced bio networks → better performed deep networks?
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