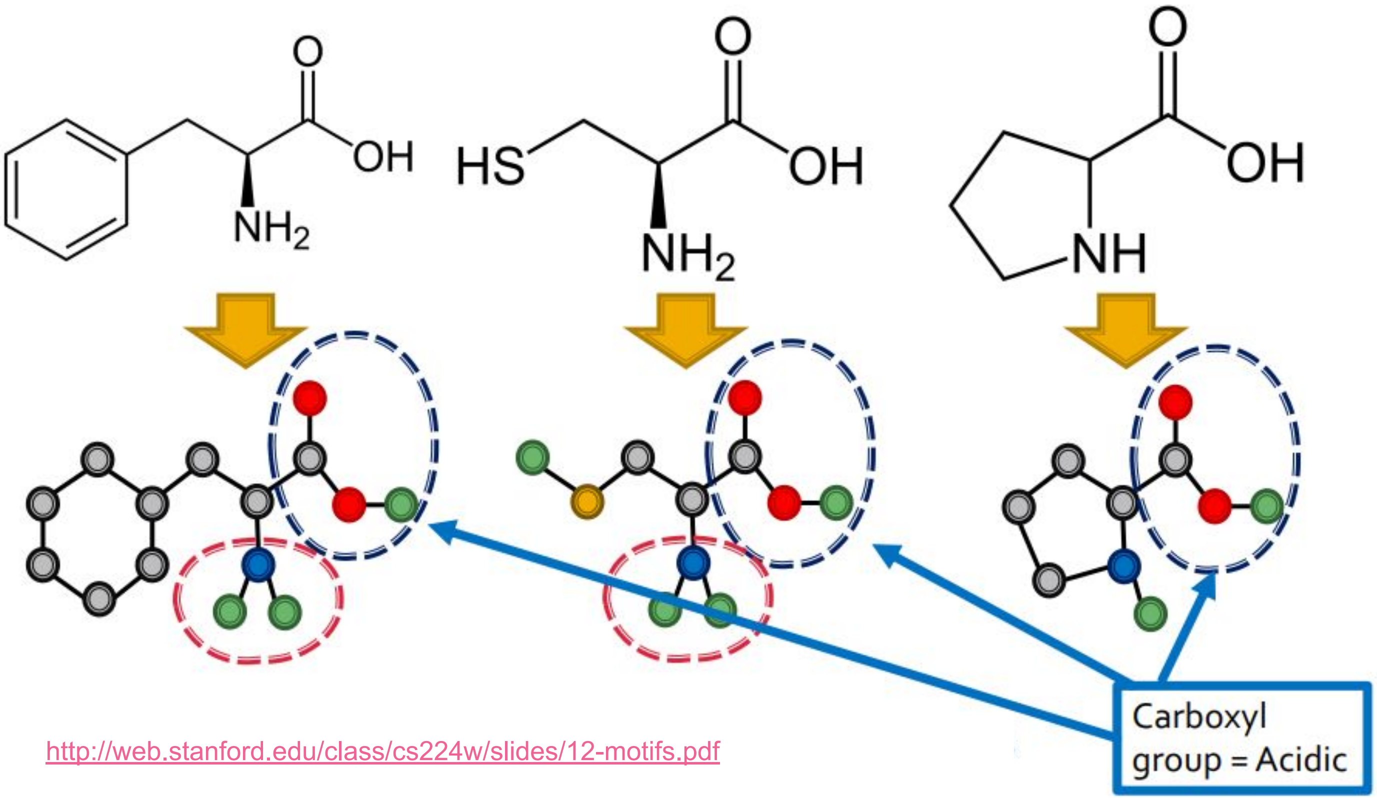
# Final presentation

Nicolas Alder, Maximilian Kleissl, Til Schniese

# Problem




http://web.stanford.edu/class/cs224w/slides/12-motifs.pdf

# Problem

Does $G$ contain a complete subgraph with $k$ vertices?

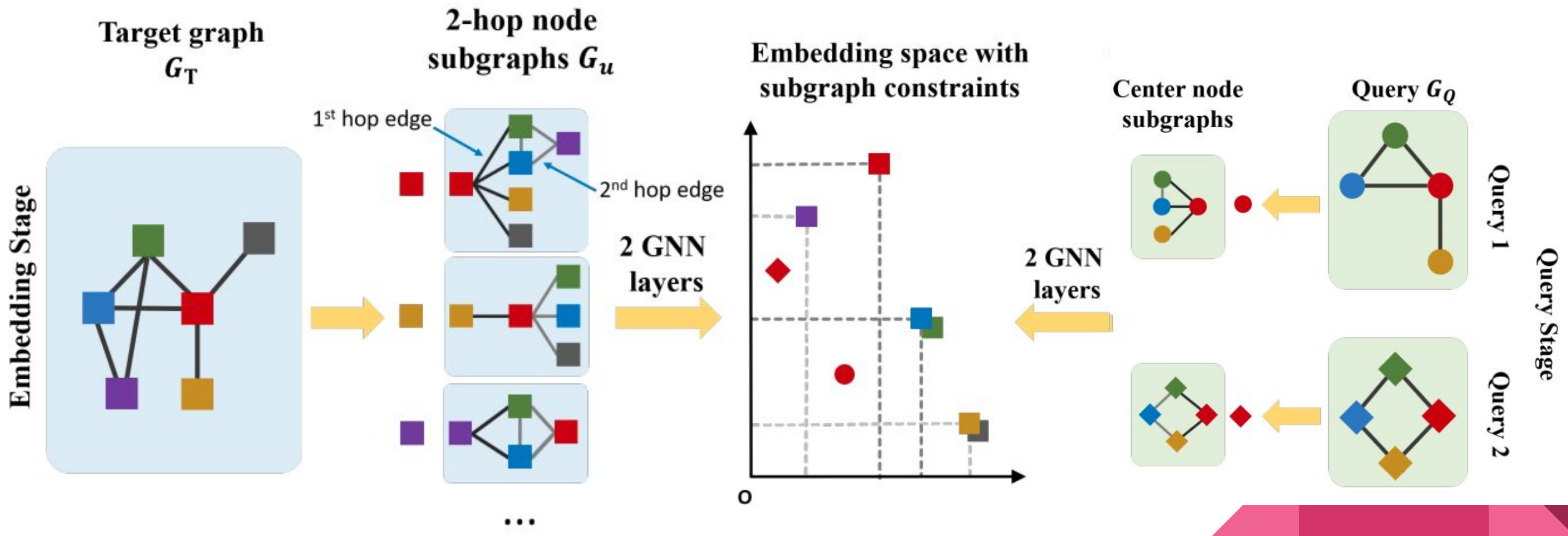
=

Does $G$ contain a subgraph $H$, with $H$ being a complete graph with $k$ vertices?

# Table of Contents

# NeuroMatch

# NeuroMatch - Architecture



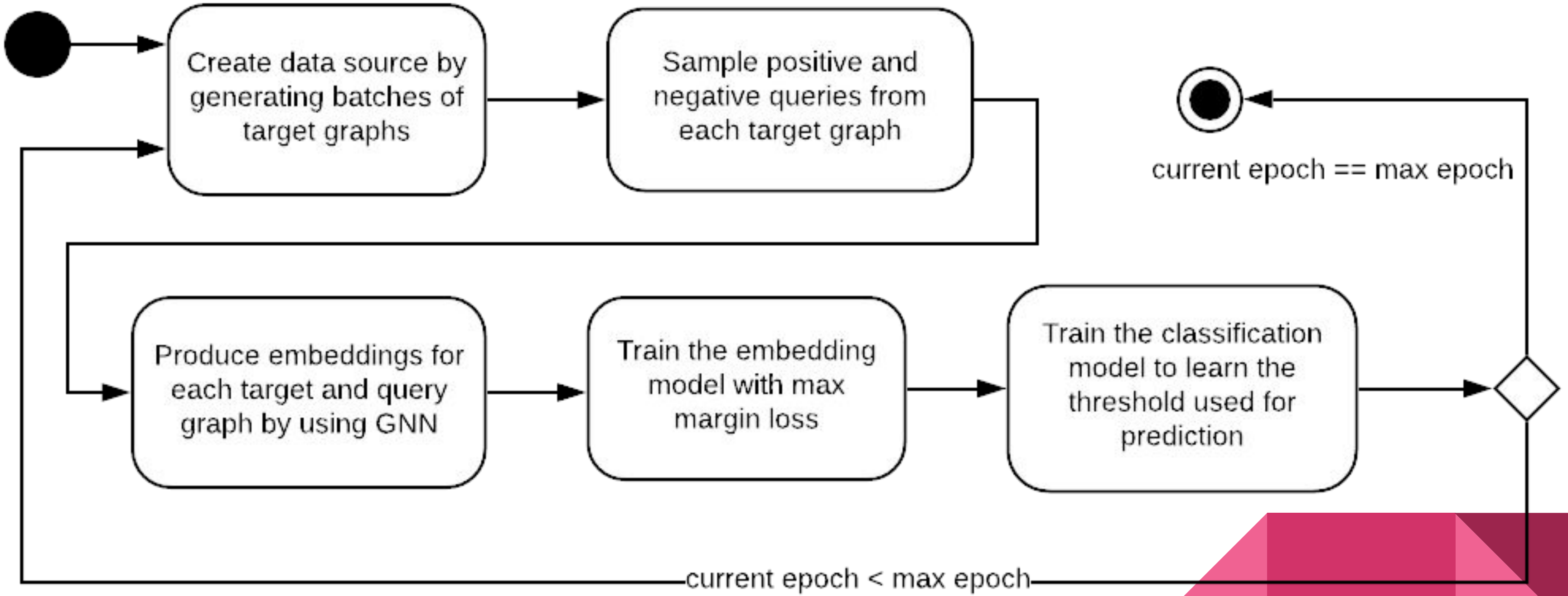
# NeuroMatch - Prediction Function

$$\mathcal{L}(z_q, z_u) = \sum_{(z_q, z_u) \in P} E(z_q, z_u) + \sum_{(z_q, z_u) \in N} \max\{0, \alpha - E(z_q, z_u)\}, \text{where}$$

$$E(z_q, z_u) = ||\max\{0, z_q - z_u\}||_2^2$$

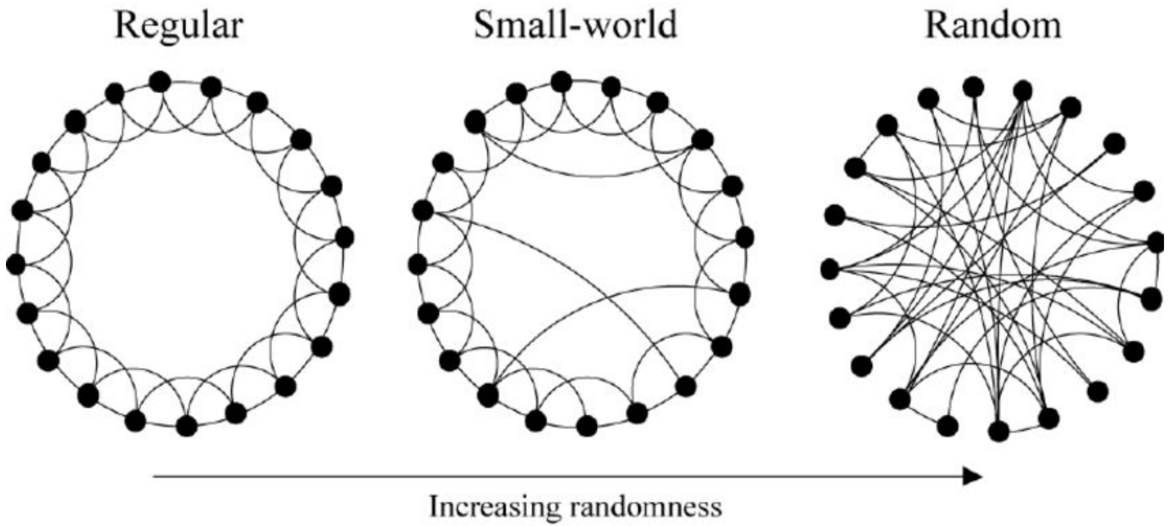
$$f(z_q, z_u) = \begin{cases} 1 & \text{iff } E(z_q, z_u) < t \\ 0 & \text{otherwise} \end{cases}$$

# NeuroMatch - Sampling & Training



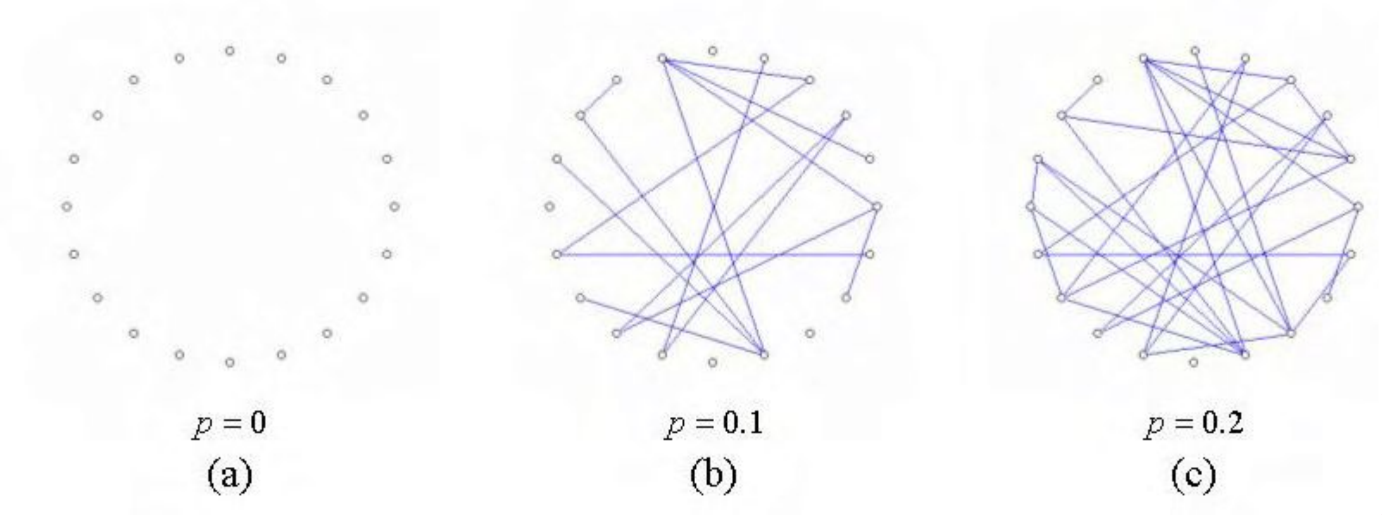
# Concepts - Dataset generators (1)

Connected Watts Strogatz graph (WS)



# Concepts - Dataset generators (1)

Erdős–Rényi graph (ER)



# Initial Observations

# Research Question

**Observation**

The type of generated graph is important to converge

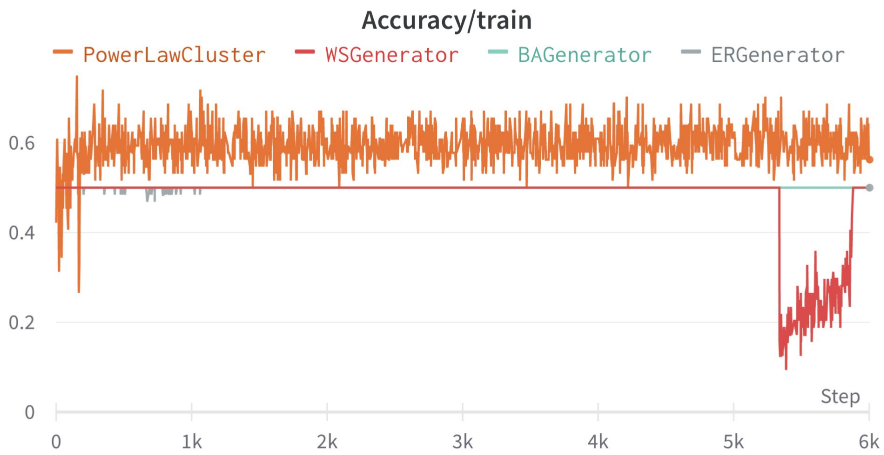
**Research Question**

Does the performance differ when limiting the dataset to a specific generator?

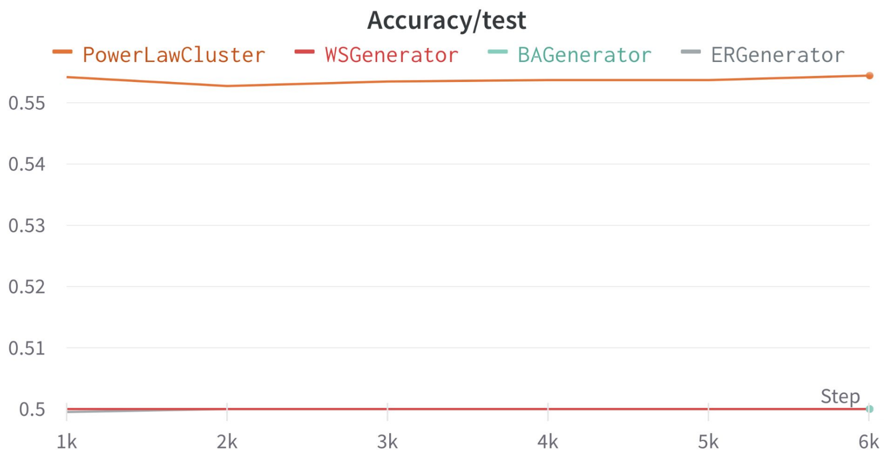
**Research Objective**

Train model on each generator in isolation

# Results



Accuracy/train
PowerLawCluster
WSGenerator
BAGenerator
ERGenerator
0.6
0.4
0.2
0
Step
0
1k
2k
3k
4k
5k
6k

Accuracy/test
PowerLawCluster
WSGenerator
BAGenerator
ERGenerator
0.55
0.54
0.53
0.52
0.51
0.5
Step
1k
2k
3k
4k
5k
6k

# Experiments

# Research Question

**Observation**
The structural properties of graphs have influence on the learning process.

**Research Question**
Which structural properties might have an effect on the learning process?

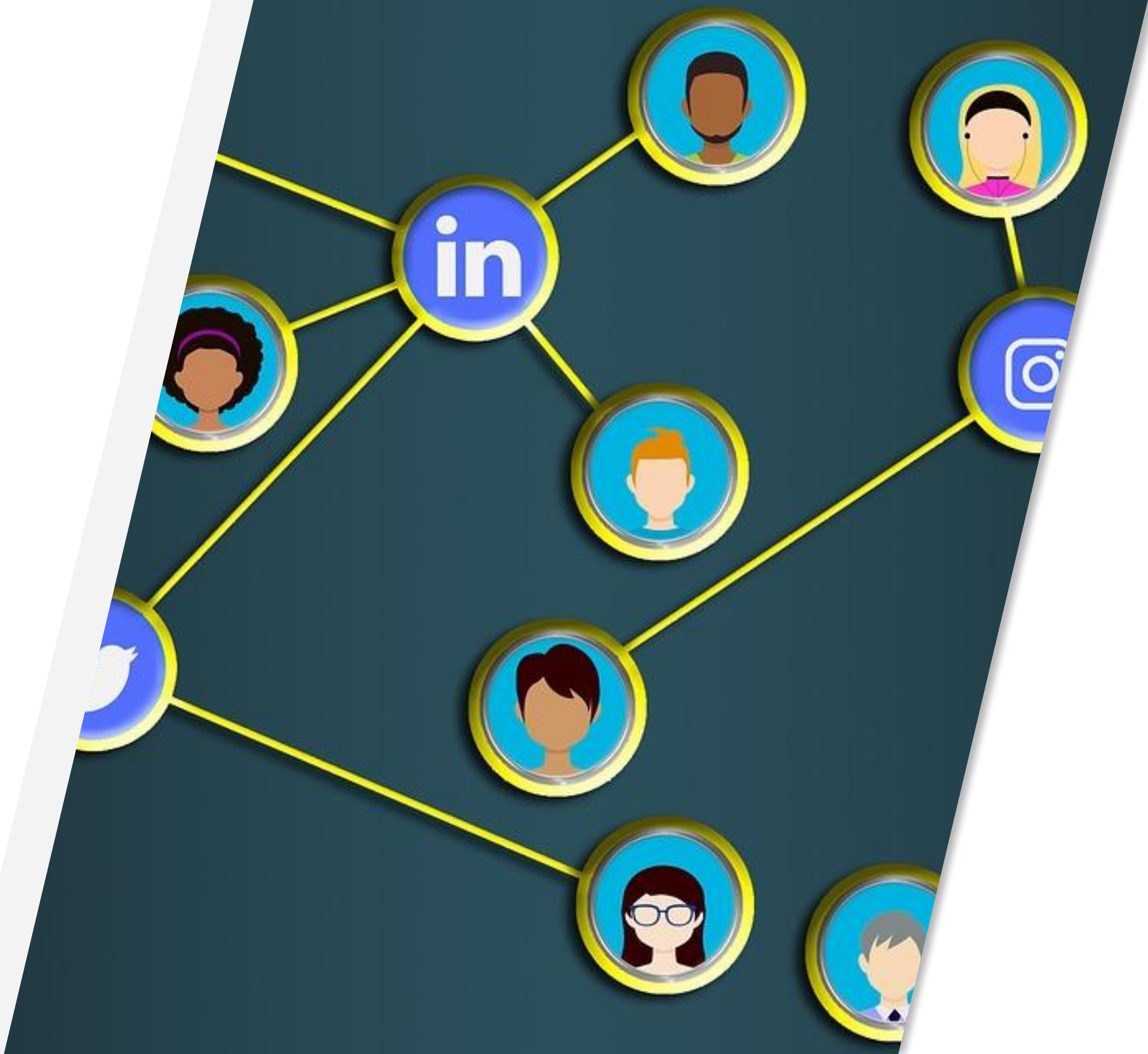
**Research Objective**
Examine structural shifts in #nodes, #edges, eccentricity, target graph size on training accuracy.

# Experiment Setup

| | Our Experiment | NeuroMatch |
|---|---|---|
| **Target Graph Generation** | once | every epoch |
| **Query Graph Generation** | every epoch | every epoch |
| **Target Graph Sizes** | one size from [11, 18] | mix from [5, 29] |
| **Shift Parameter** | [0.55, 0.71, 0.83, 1.0, 1.2, 1.6, 1.8] | 1.0 |

# Concepts - Graph Metrics

- Number of nodes
- Number of edges
- Eccentricity
    - minimal: radius
    - maximal: diameter



# Modifiying Erdős–Rényi graphs

Original

```
mean = log2(num_nodes) / num_nodes

beta = alpha / mean - alpha

p = random.beta(alpha, beta)

graph = gnp_random_graph(num_nodes, p)
```

Change number of edges

```
mean = (log2(num_nodes) / num_nodes) * shift_parameter

beta = alpha / mean - alpha

p = random.beta(alpha, beta)

graph = gnp_random_graph(num_nodes, p)
```

# Modifiying Erdős–Rényi graphs

## Original

```
mean = log2(num_nodes) / num_nodes

beta = alpha / mean - alpha

p = random.beta(alpha, beta)

graph = gnp_random_graph(num_nodes, p)
```

## Change number of nodes

```
num_nodes = num_nodes * shift_parameter

mean = (log2(num_nodes) / num_nodes) *
edge_rescale

beta = alpha / mean - alpha

p = random.beta(alpha, beta)

graph = gnp_random_graph(num_nodes, p)
```

# Modifiying Watts-Strogatz graphs

Original

```
density_mean = log2(num_nodes) / num_nodes

density_beta = density_alpha / density_mean - density_alpha

k = random.beta(density_alpha, density_beta) * num_nodes

graph = connected_watts_strogatz_graph(num_nodes, k, p)
```

# Modifiying Watts-Strogatz graphs

Change eccentricity (and also number of nodes and number of edges)

```
density_mean = log2(num_nodes) / num_nodes * shift_parameter

density_beta = density_alpha / density_mean - density_alpha

k = random.beta(density_alpha, density_beta) * num_nodes

graph = connected_watts_strogatz_graph(num_nodes, k, p)
```

# Findings

# Planned Experiments = 720

(8 target graph sizes * 10 shift parameters * 3 generators * 3 experiment runs)

# Successful Experiments = 612

Result Datasets: ER Nodes, ER Edges, WS Eccentricity, all data combined

Two regression analysis for each result dataset:

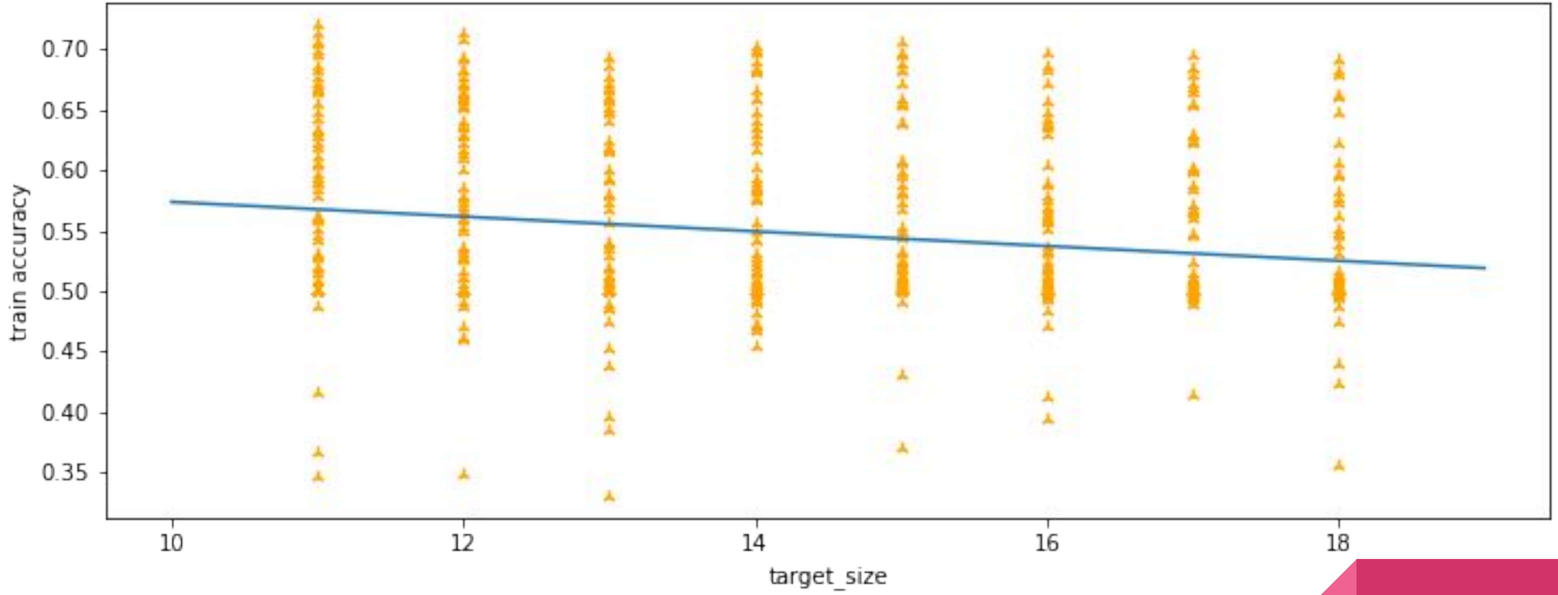
- Does the target graph size has influence on training accuracy?
  (independent: target graph size, dependent: training accuracy)

- Does the shift parameter has influence on training accuracy?
  (independent: shift parameter, dependent: training accuracy)

# Findings - Target Graph Sizes

| | Adj. R-Squared | Coefficient | P-Value | Sample Size |
|---|---|---|---|---|
| ER Nodes | 0.103 | -0.0060 | 0.000 | 183 |
| ER Edges | 0.067 | -0.0072 | 0.000 | 213 |
| WS Eccentricity | 0.015 | -0.0046 | 0.039 | 216 |
| All Generators | 0.037 | -0.0061 | 0.000 | 612 |

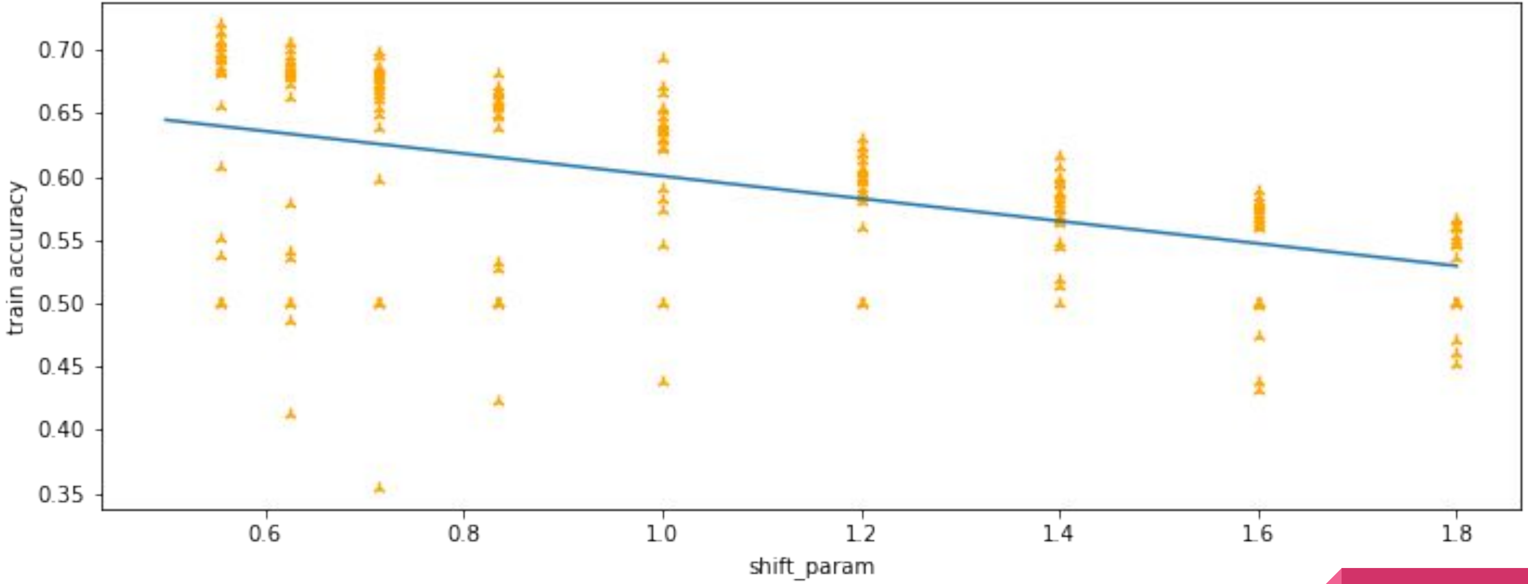
# Findings - Target Graph Sizes



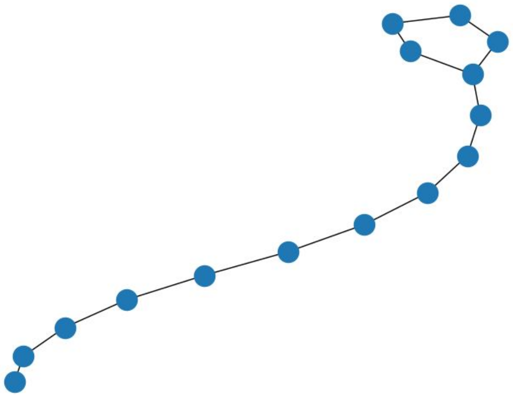
# Findings - Shift Parameter

| | Adj. R-Squared | Effect Size / Coefficient | P-Value | Sample Size |
|---|---|---|---|---|
| ER Nodes | 0.006 | -0.0115 | 0.144 | 183 |
| ER Edges | 0.066 | -0.0390 | 0.000 | 213 |
| WS Eccentricity | 0.242 | -0.0887 | 0.000 | 216 |
| All Generators | 0.099 | -0.0545 | 0.000 | 612 |

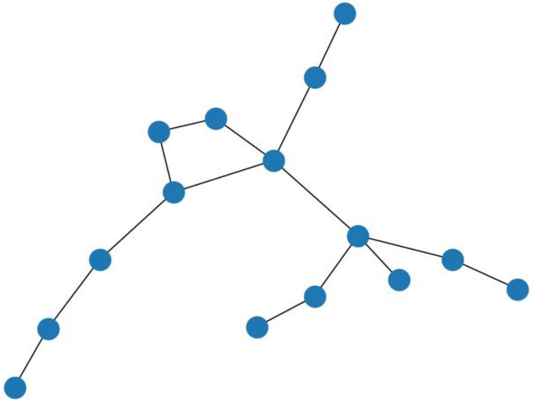
# Findings - Shift Parameter (WS Eccentricity)



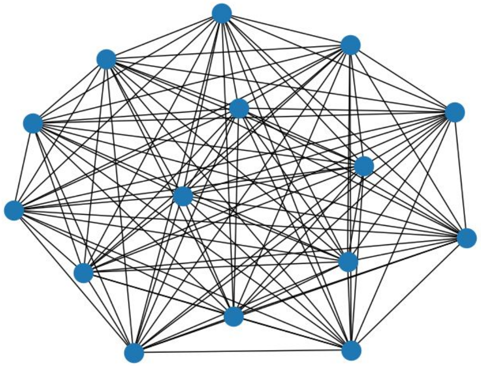
# Findings - Low vs. High Eccentricity



(size = 15, shift = 0.55)

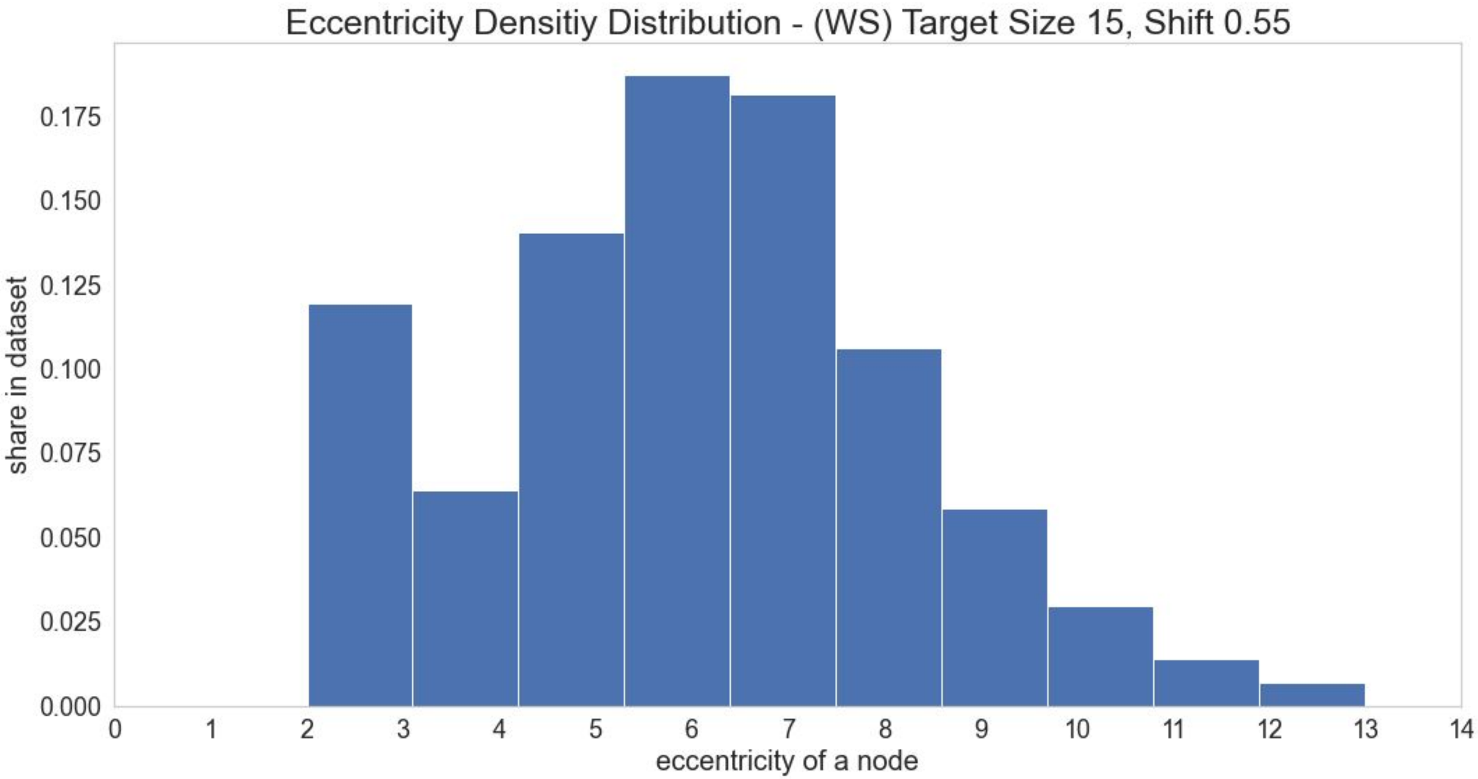


(size = 15, shift = 1.00)

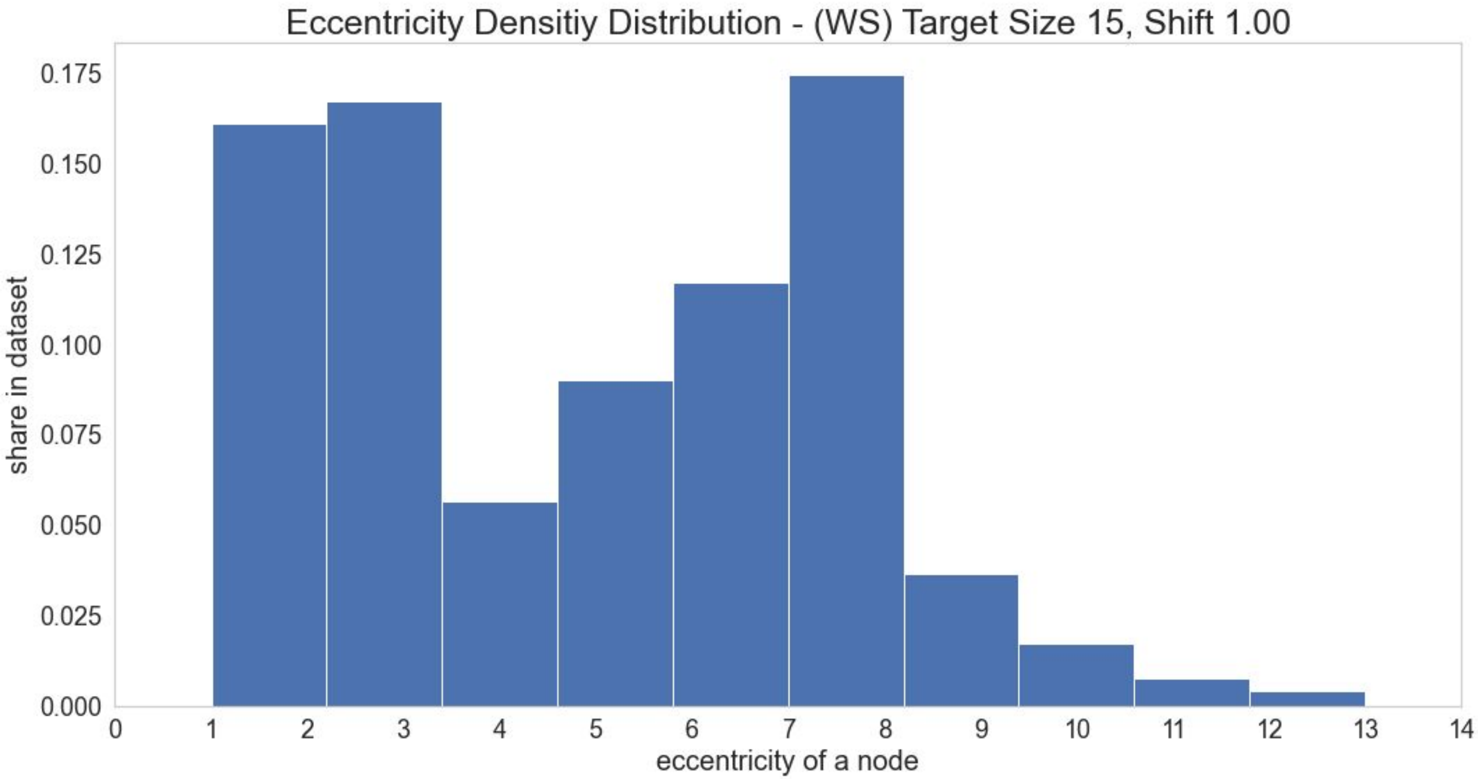


(size = 15, shift = 1.8)

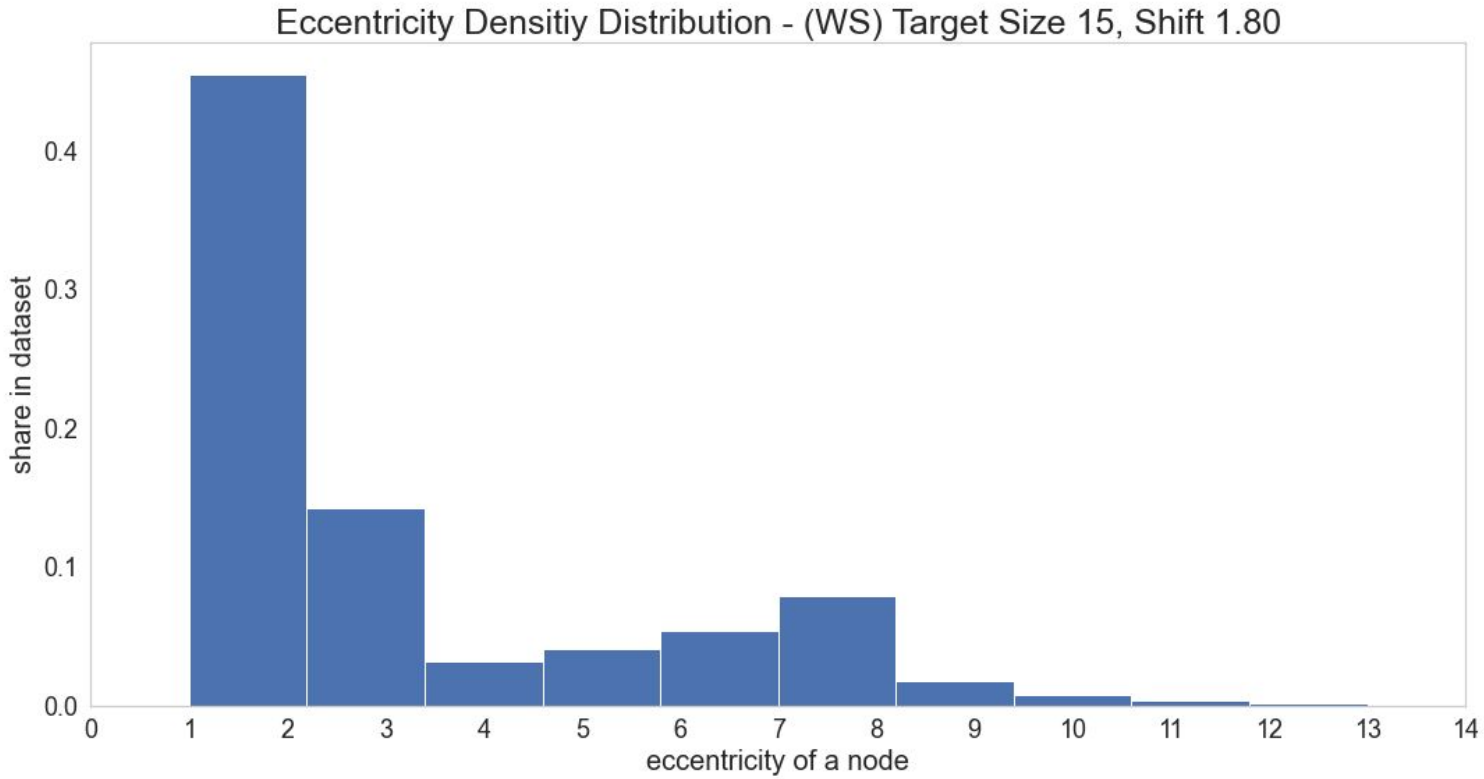
# Findings - Low vs. High Eccentricity



# Findings - Low vs. High Eccentricity



# Findings - Low vs. High Eccentricity



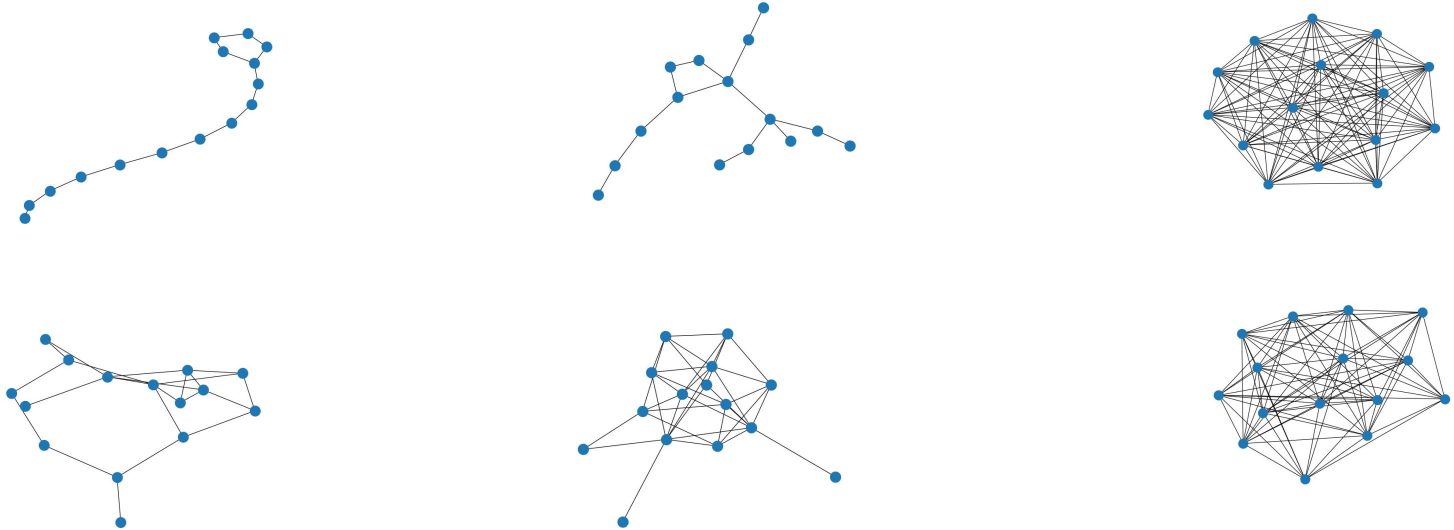
# Findings - Low vs. High Eccentricity

| **(WS) Shift** | **0.55** | **1.0** | **1.8** |
|---|---|---|---|
| **Mean Eccentricity** | 6.2 | 5.0 | 3.4 |
| **Mean Radius** | 4.3 | 3.6 | 2.6 |
| **Mean Diameter** | 7.7 | 6.2 | 4.1 |

# Findings - WS vs. ER

**(WS)**

**(ER Edges)**



**(size = 15, shift = 0.55)** **(size = 15, shift = 1.00)** **(size = 15, shift = 1.8)**

# Findings - Shift Parameter

| | Adj. R-Squared | Effect Size / Coefficient | P-Value | Sample Size |
|---|---|---|---|---|
| ER Nodes | 0.006 | -0.0115 | 0.144 | 183 |
| ER Edges | 0.066 | -0.0390 | 0.000 | 213 |
| WS Eccentricity | 0.242 | -0.0887 | 0.000 | 216 |
| All Generators | 0.099 | -0.0545 | 0.000 | 612 |

# Threats to validity

- Experiment repetitions limited
- Experiment duration limited
- Query:target graph size ratio fixed
- Limited expressiveness for heterogeneous composited datasets
  - how do real-world datasets look like
  - generalization to real-world dataset
- Might not be the main driver of learning
  - why do real-world datasets are reported easier than synthetic → analyze
- NeuroMatch results are hardly reproducible
- Limitation to NeuroMatch approach

# Conclusion and Future Work

- Eccentricity has influence on performance
- We do not know if this is actually the main driver for learning (probably not, more like a factor)
- Graph structure is more than simple number of nodes/edges

Future Work:

- Real-World Datasets
- Influence of the specific NeuroMatch approach on learning characteristics