New in ML

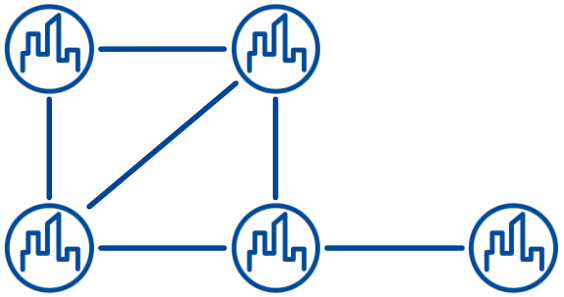
Edge Contraction Pooling for Graph Neural Networks

Frederik Diehl

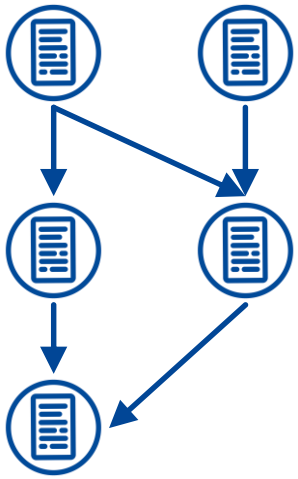
Bobak HaEri

Many Real-World Problems are non-Euclidian

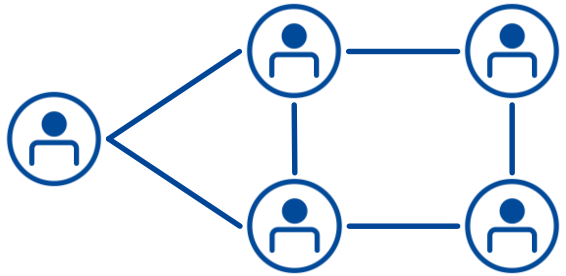
Road Graphs



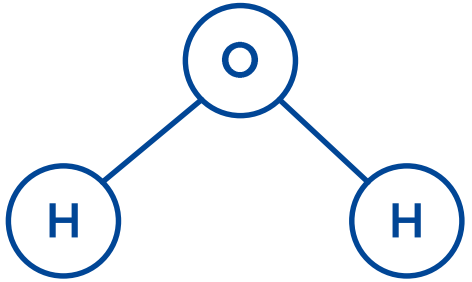
Citation Graphs



Social Graphs

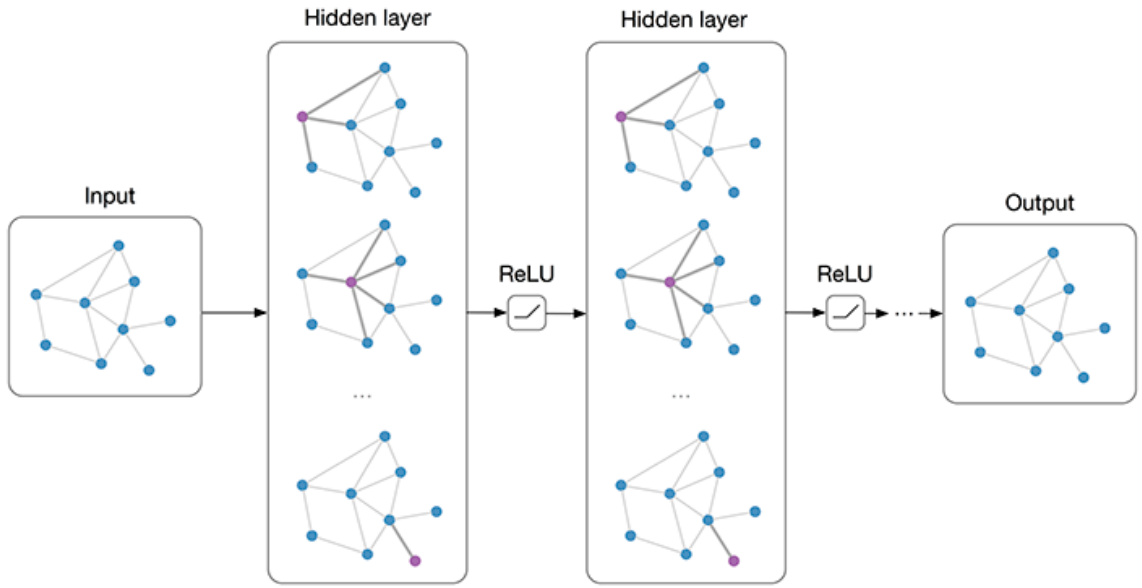


Molecules



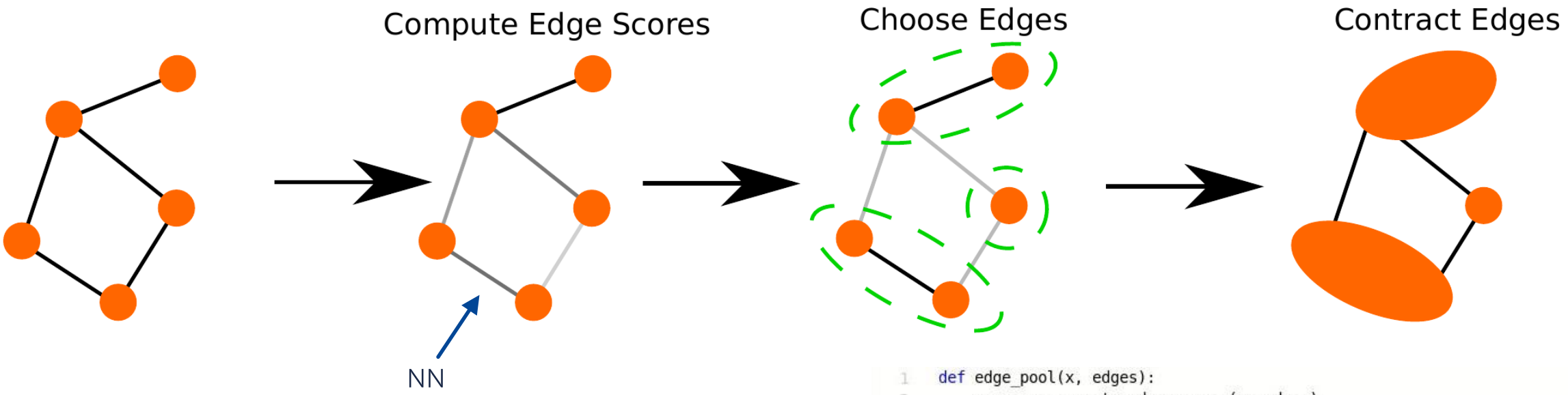
Graph Neural Networks

- Generalize convolutional networks to non-Euclidian data
- Usually based on a message-passing framework



tkipf.github.io/graph-convolutional-networks/

EdgePool




```python
def edge_pool(x, edges):
    scores = compute_edge_scores(x, edges)
    edges = sort_edges_by_score(edges, scores)
    expanded = set()
    for edge in edges:
        if not (edge.node_from in expanded or edge.node_to in expanded):
            expanded.add(edge)

    new_edges = expanded
    new_x = compute_node_features(x, edges)
    return new_x, new_edges
```

Code: Available in pytorch-geometric

EdgePool – Advantages and Limitations

Advantages

- Local, hard pooling
- Sparse
- Scales in O(e log(e))

Limitations

- Slow and hard to parallelize
- Restricted to pair-wise pooling

Results

EdgePool Results

Graph Classification

	PROTEINS	RDT-B	RDT-12K	COLLAB
Base Model	71.4 ± 3.2	69.9 ± 3.7	35.1 ± 1.6	65.4 ± 1.5
DiffPool [*]	72.3 ± 5.8	82.9 ± 3.4	34.8 ± 1.9	**70.1 ± 1.5**
TopKPool	70.6 ± 4.8	68.9 ± 3.2	28.7 ± 1.8	64.6 ± 2.1
SAGPool	71.8 ± 6.0	84.7 ± 4.4	41.9 ± 3.3	63.9 ± 2.5
EdgePool	**72.5 ± 3.2**	**87.3 ± 4.1**	**45.6 ± 1.8**	67.1 ± 2.7

+2 p.p.

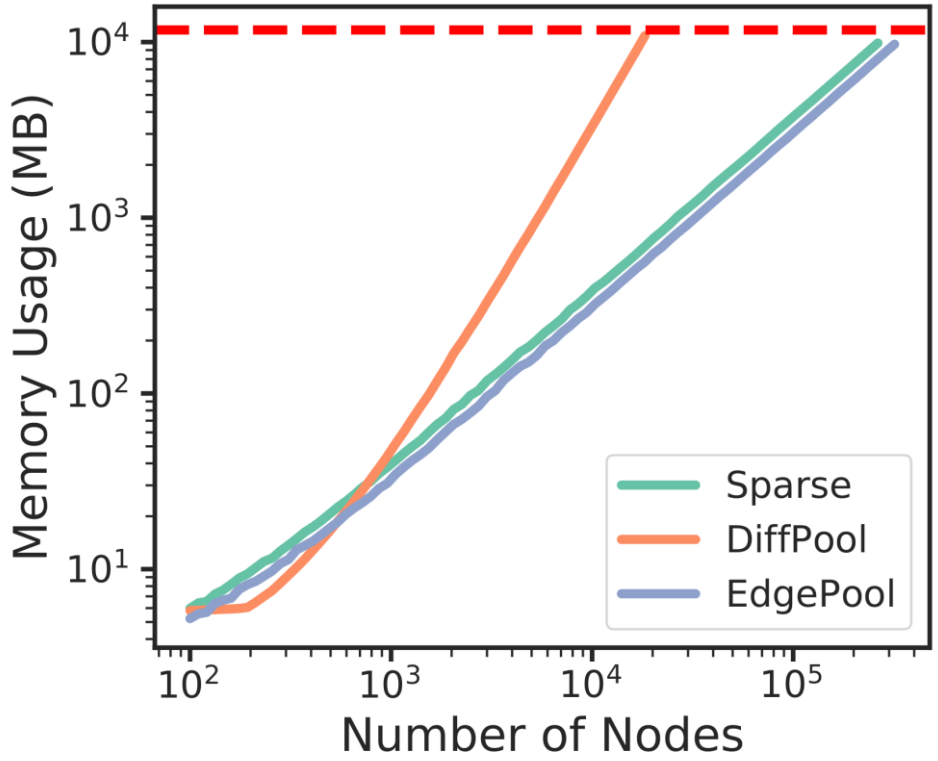
Node Classification

CORA	GCN	GIN	GIN0	GAT	MLP
No Pooling	71.8 ± 3.4	52.1 ± 4.7	55.9 ± 4.4	68.0 ± 4.5	35.6 ± 2.6
EdgePool	**72.8 ± 1.9**	**63.0 ± 5.4**	**61.3 ± 3.9**	**70.3 ± 3.3**	**58.3 ± 3.6**
CITESEER					
No Pooling	62.9 ± 2.9	40.9 ± 4.6	41.4 ± 3.8	58.9 ± 2.8	35.5 ± 3.2
EdgePool	**65.3 ± 2.7**	**50.6 ± 3.9**	**49.9 ± 5.7**	**61.0 ± 3.4**	**50.0 ± 3.7**
PUBMED					
No Pooling	**74.2 ± 1.7**	60.8 ± 6.8	61.0 ± 4.4	**73.0 ± 2.0**	62.4 ± 4.1
EdgePool	74.1 ± 2.1	**61.0 ± 6.4**	**61.9 ± 4.9**	72.0 ± 4.7	**64.8 ± 3.2**
PHOTO					
No Pooling	**88.4 ± 2.2**	69.9 ± 3.2	71.9 ± 4.0	78.5 ± 4.5	59.6 ± 4.9
EdgePool	86.5 ± 0.8	**77.1 ± 1.8**	**78.1 ± 1.5**	**81.0 ± 4.2**	**81.4 ± 2.3**
COMPUTER					
No Pooling	**80.0 ± 2.6**	53.1 ± 5.5	52.4 ± 3.6	60.6 ± 12.4	43.0 ± 6.7
EdgePool	77.9 ± 2.2	**58.1 ± 4.8**	**60.4 ± 4.3**	**62.5 ± 13.0**	**69.4 ± 2.3**

+3.5 p.p.

EdgePool improves performance on many models and datasets

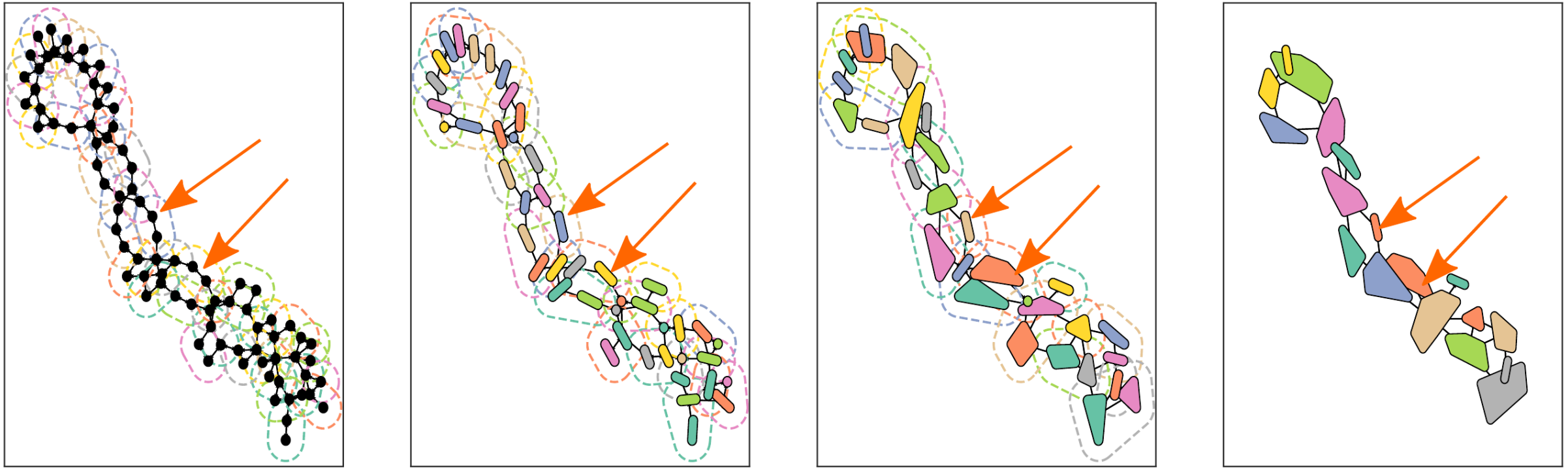
Memory Consumption

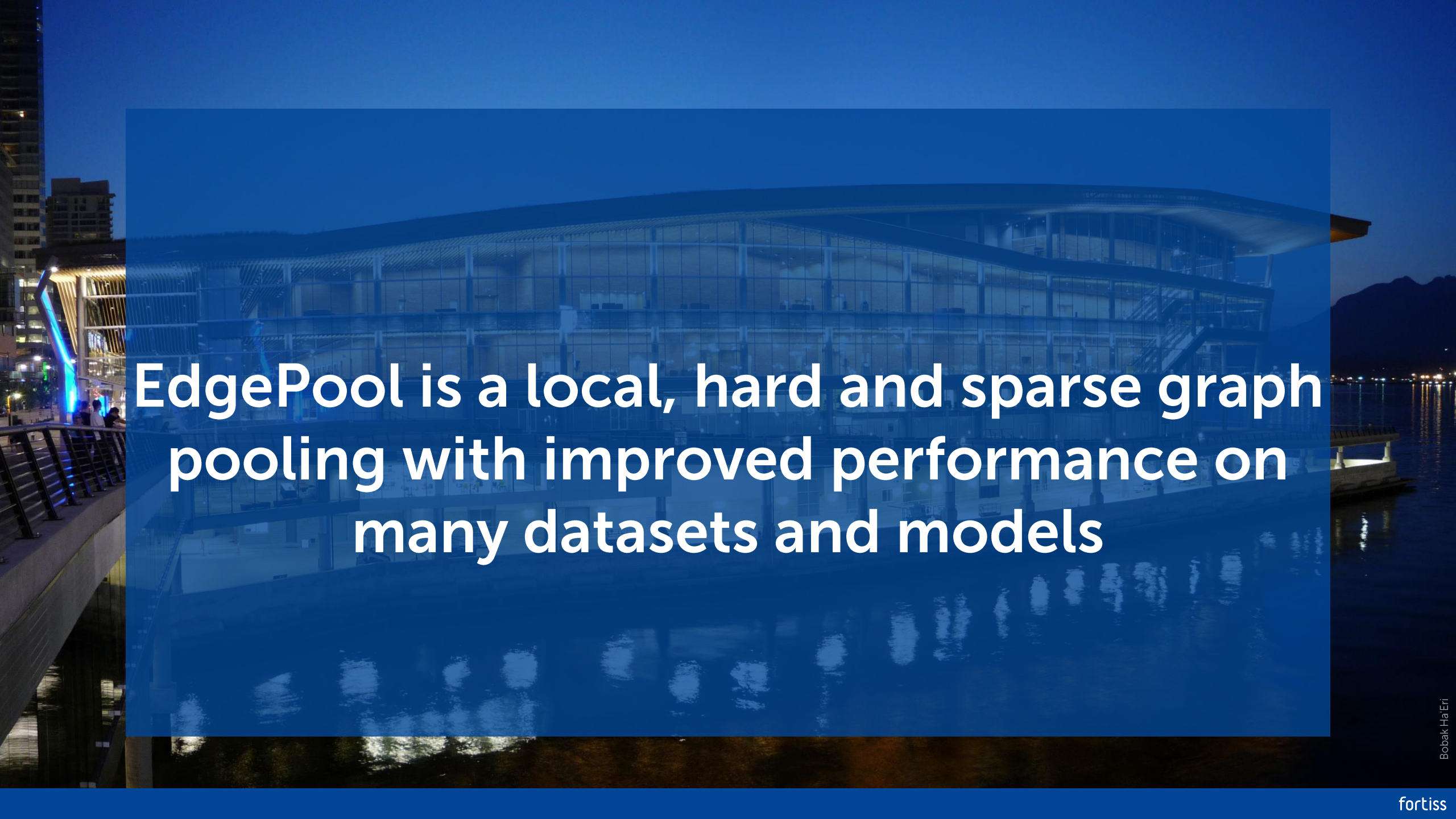


Limits

- DiffPool: 18k
- Sparse: 250k
- EdgePool: 300k

EdgePool Visualization





EdgePool is a local, hard and sparse graph pooling with improved performance on many datasets and models

Bobak Ha'Eri

Thank you.

Questions?

Bobak Ha'Eri

Contact

fortiss GmbH
Guerickestraße 25
80805 Munich
GERMANY
www.fortiss.org
info@fortiss.org