東京工業大学
Tokyo Institute of Technology

MURATA LAB

Research Progress

Motif-aware method for graph analysis

Hoang Nguyen (M2), Supervisor: Assoc. Prof. Tsuyoshi Murata

2017/04/13

Roadmap

Result of motif methods | Proposal of future models

NETWORK MOTIF

Definition

Applications

Motif Conductance

EMBEDDINGS

Deep architecture

Skip-gram methods

Spectral methods

OBSERVATION

Experiments

Proposed models

Results

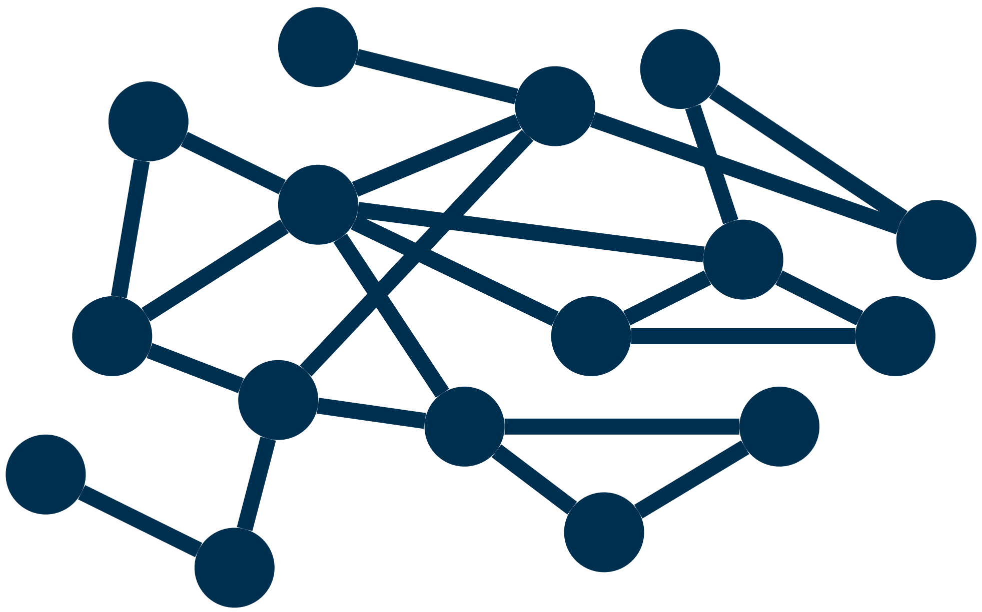
IN PROGRESS

Diversity models

Generative models

Motif Spectral

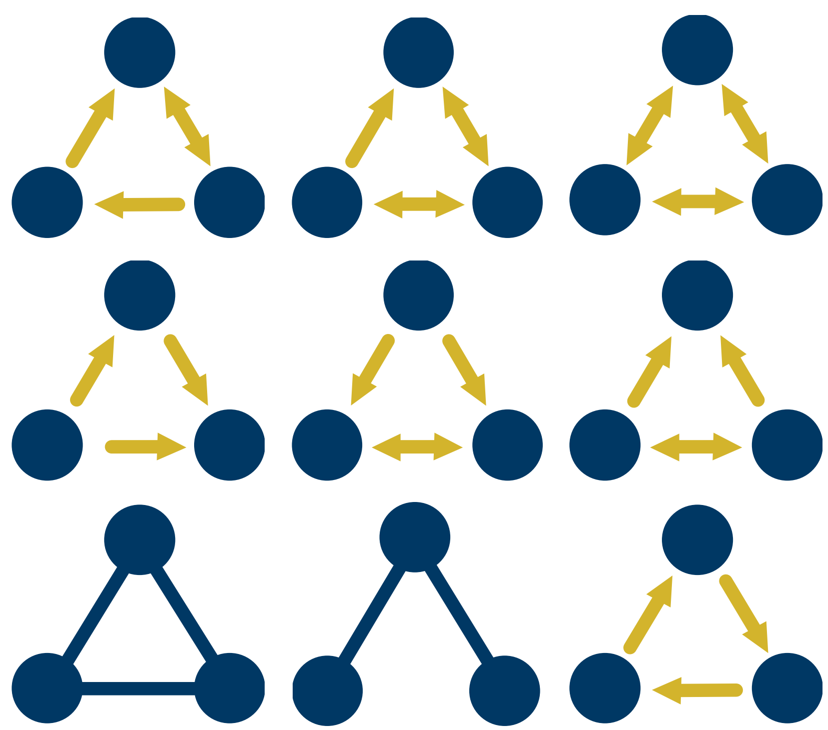
Scales of network analysis:



- Macroscopic (global view)
- Microscopic (interactions)
- Mesoscopic (sub-structures)

***Network motifs* are** [1]

> patterns of statistically significant interconnections occurring in complex networks.



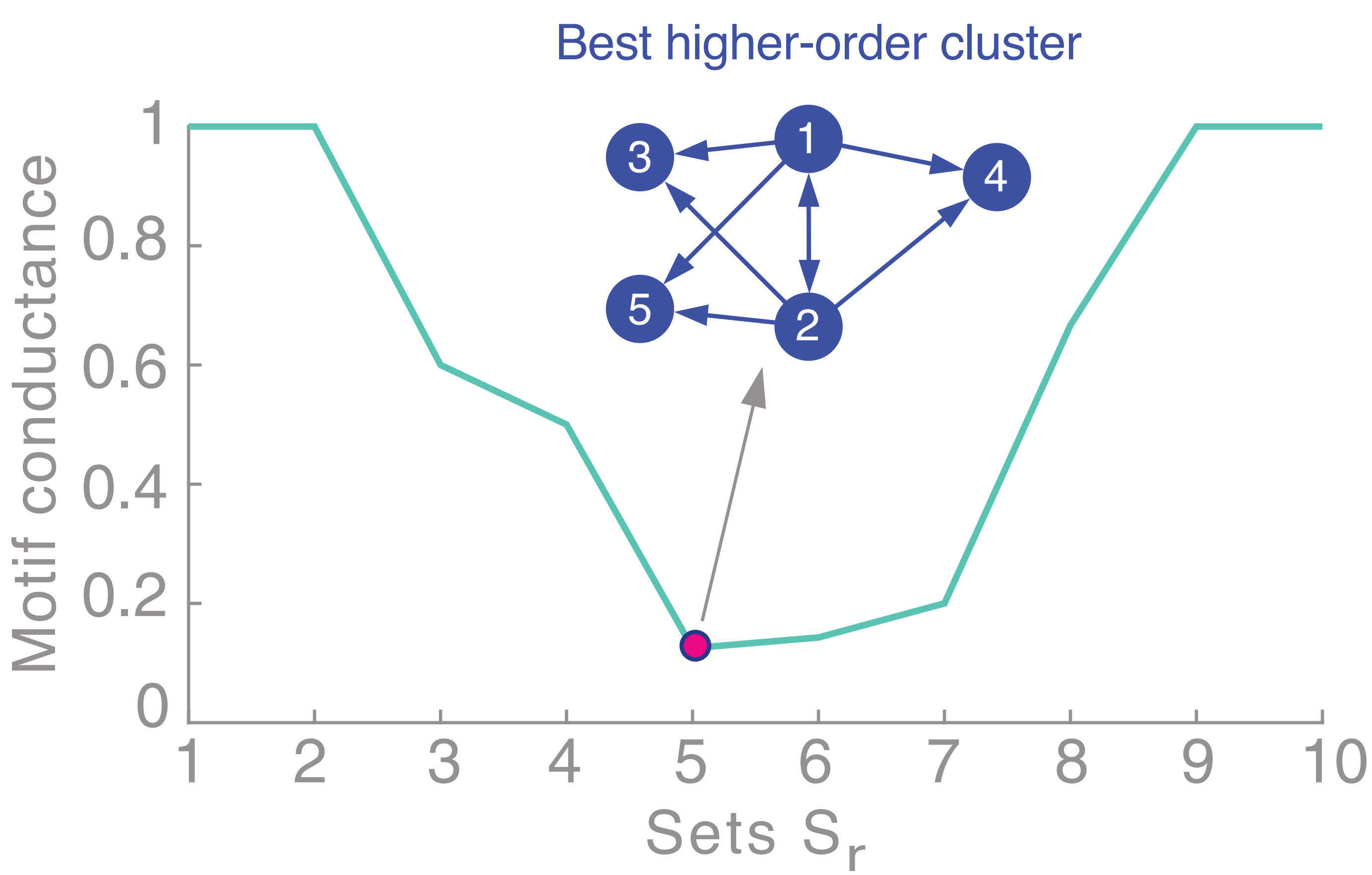
Applications

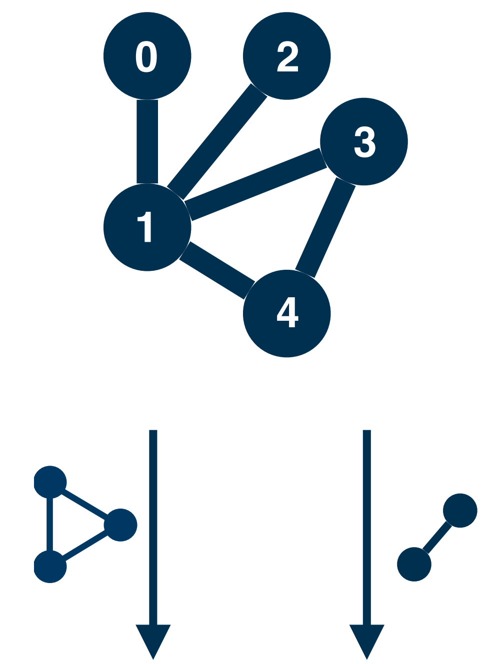
- Social networks [2]
- Biological systems [1,3,4]

[4] *(Franceschet, 2012)* [3] *(Honey, 2007)* [2] *(Holland, 1970)* [1] *(Milo, 2002)*

Motif conductance *(Benson, 2016)*

- Generalization of graph conductance.
- Motif-cooccurrence matrix.
- Good results for graph partitioning.



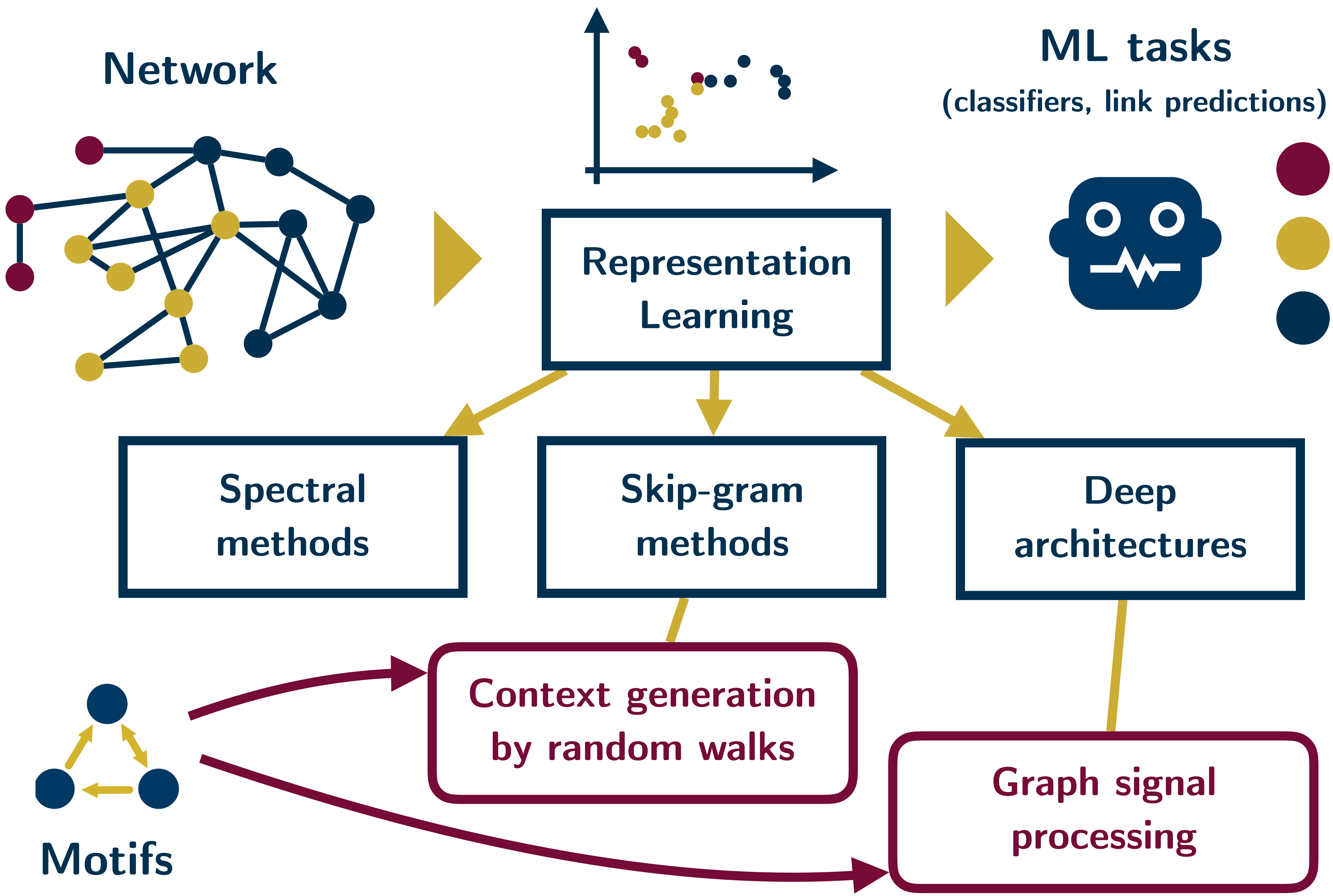


$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 2 & 2 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 2 \\ 0 & 2 & 0 & 2 & 0 \end{bmatrix}$$

$$\phi_m(S) = \frac{\mathrm{cut}_M(S, \bar{S})}{\min[\mathrm{vol}_M(S), \mathrm{vol}_M(\bar{S})]}$$

Representation Learning

Motif-aware approaches



Datasets:

- Blogcatalog3
- Cora, Citeseer, Pubmed
- Facebook (EgoNet)
- Transcription networks

Networks	***#nodes***	***#edge***
Blogcatalog3	10312	333983
Cora*	2708	5429
Citeseer*	3327	4732

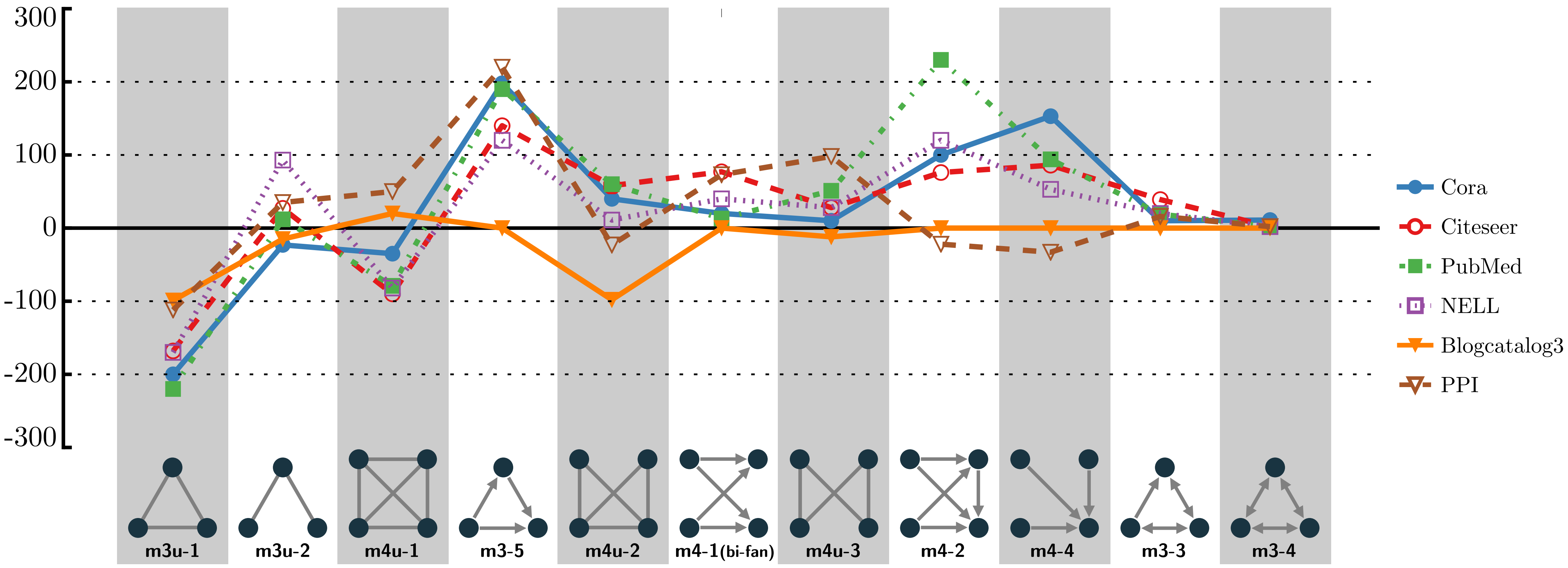
* Each node has a feature vector (tf-idf)

Deciding motif:

- Motif frequency
- Z-score
- Motif conductance

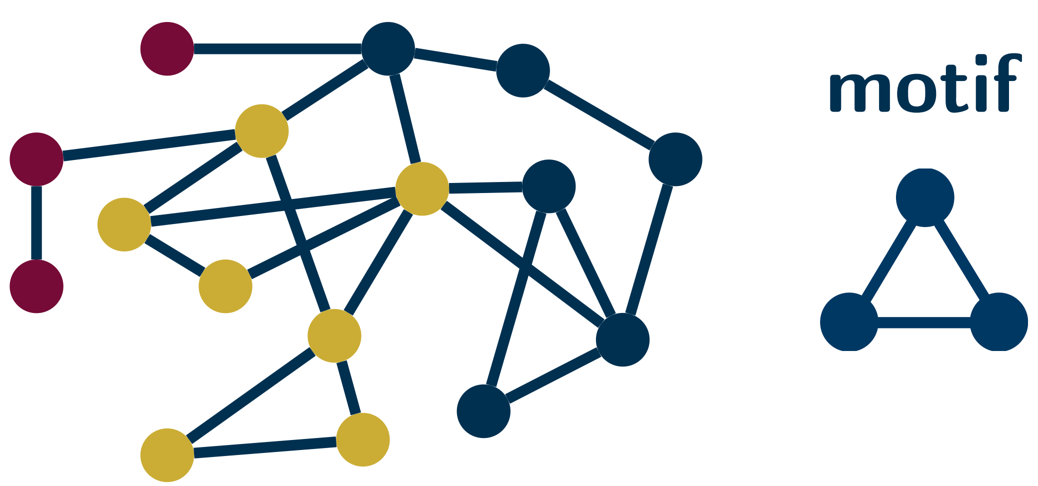
$$z_m = \frac{N_m(emp) - N_m(rnd)}{\sigma_m}$$

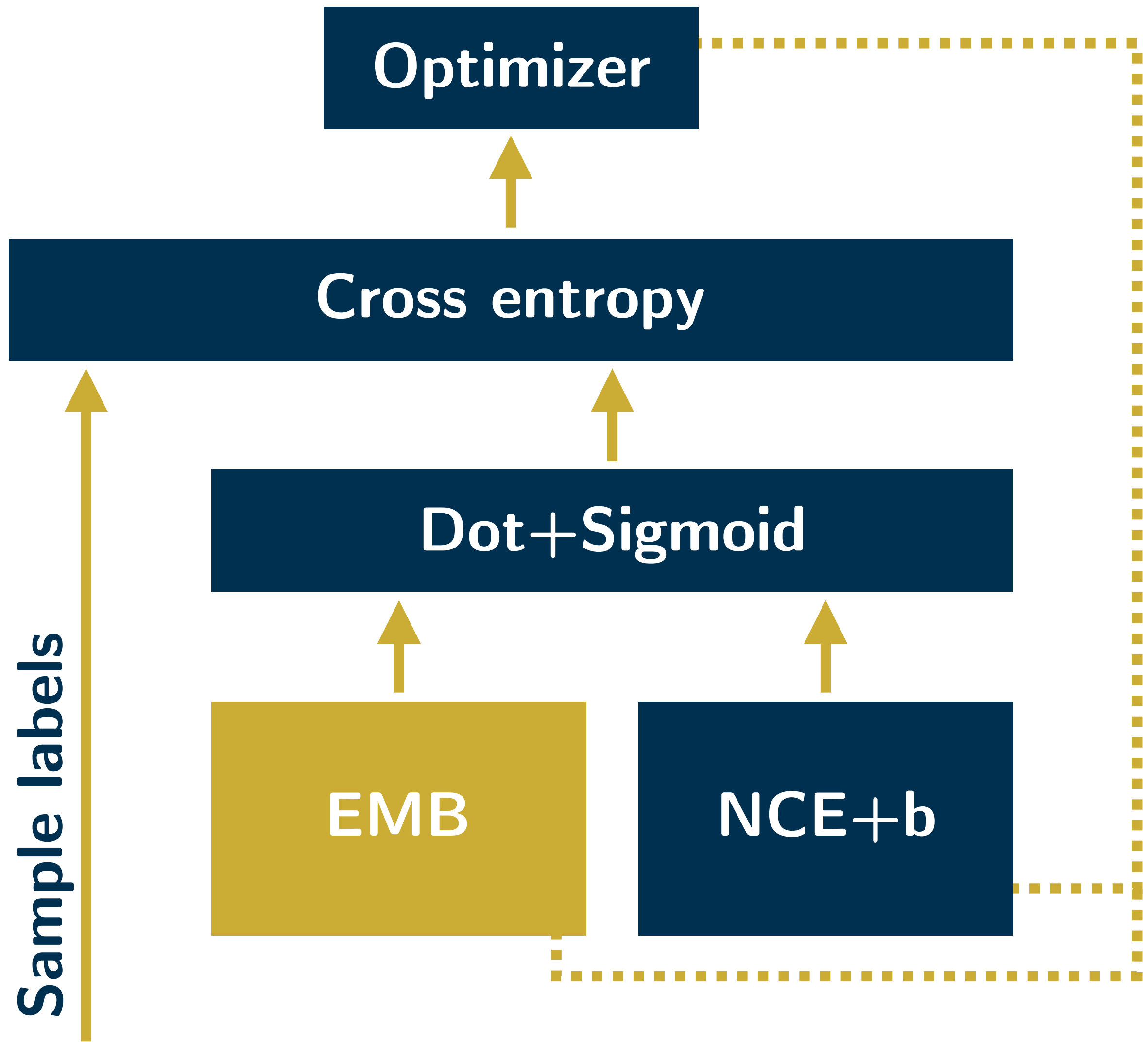
Significant graph - Selected motifs



Using motif as a guiding pattern

- DeepWalk := Skip-gram+RW
- LINE := DeepWalk+bi-fan
- node2vec := Skip-gram
  +bias-ed walk
- motifwalk := Skip-gram
  +motif walk



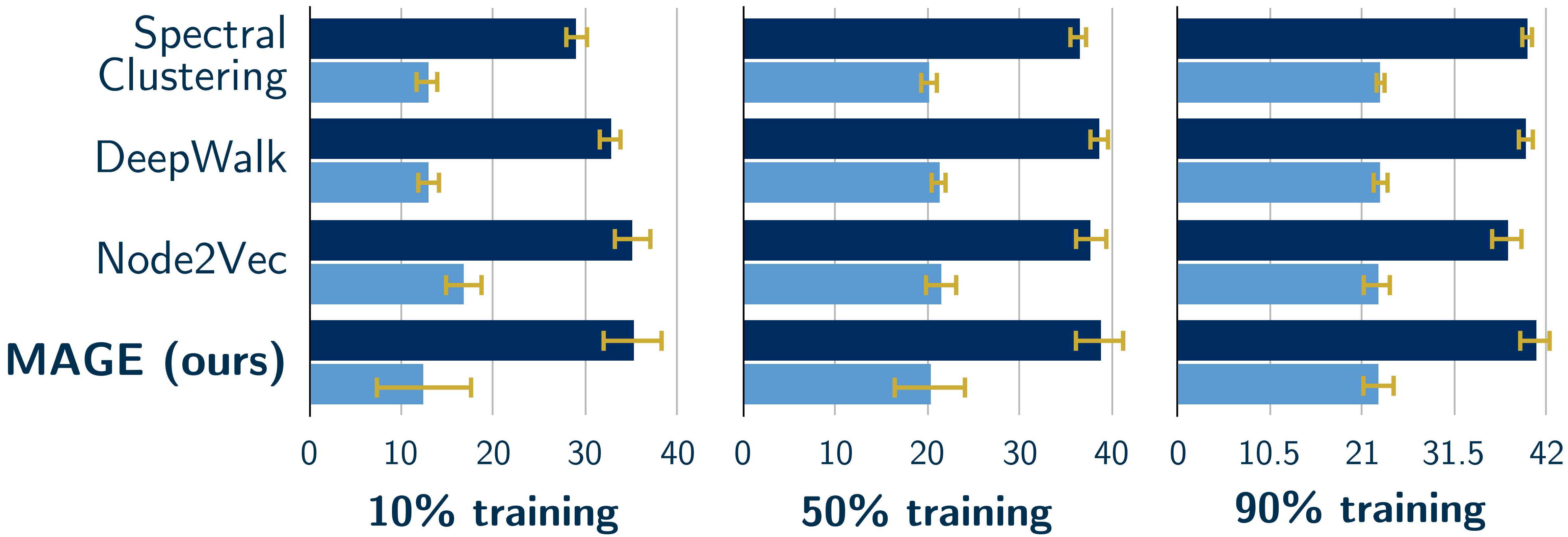


Generated context:

Motif context:

Node multi-label classification

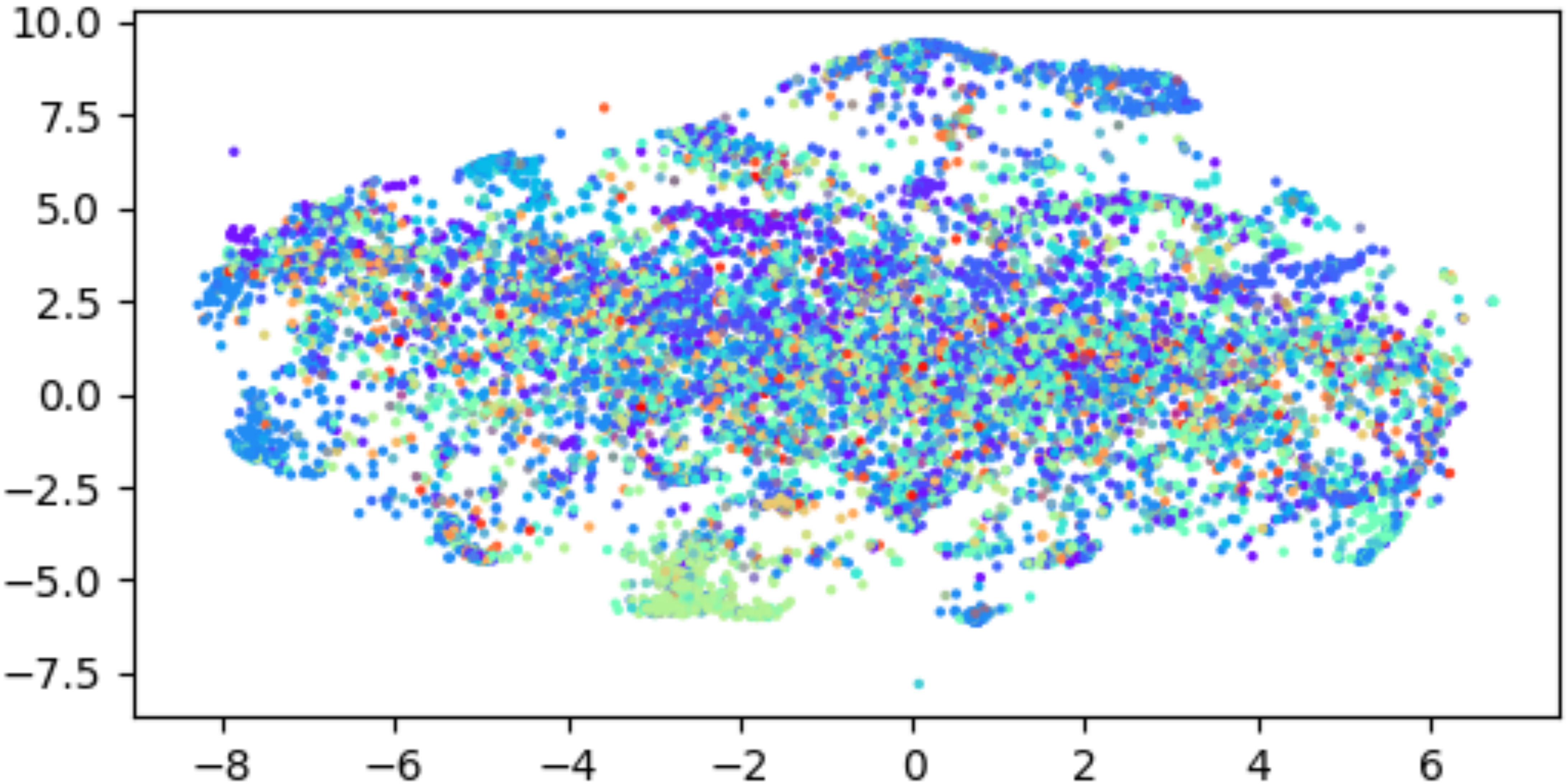
- Blogcatalog3 - f1 score
- 39 classes (imbalanced), nodes can have multiple labels



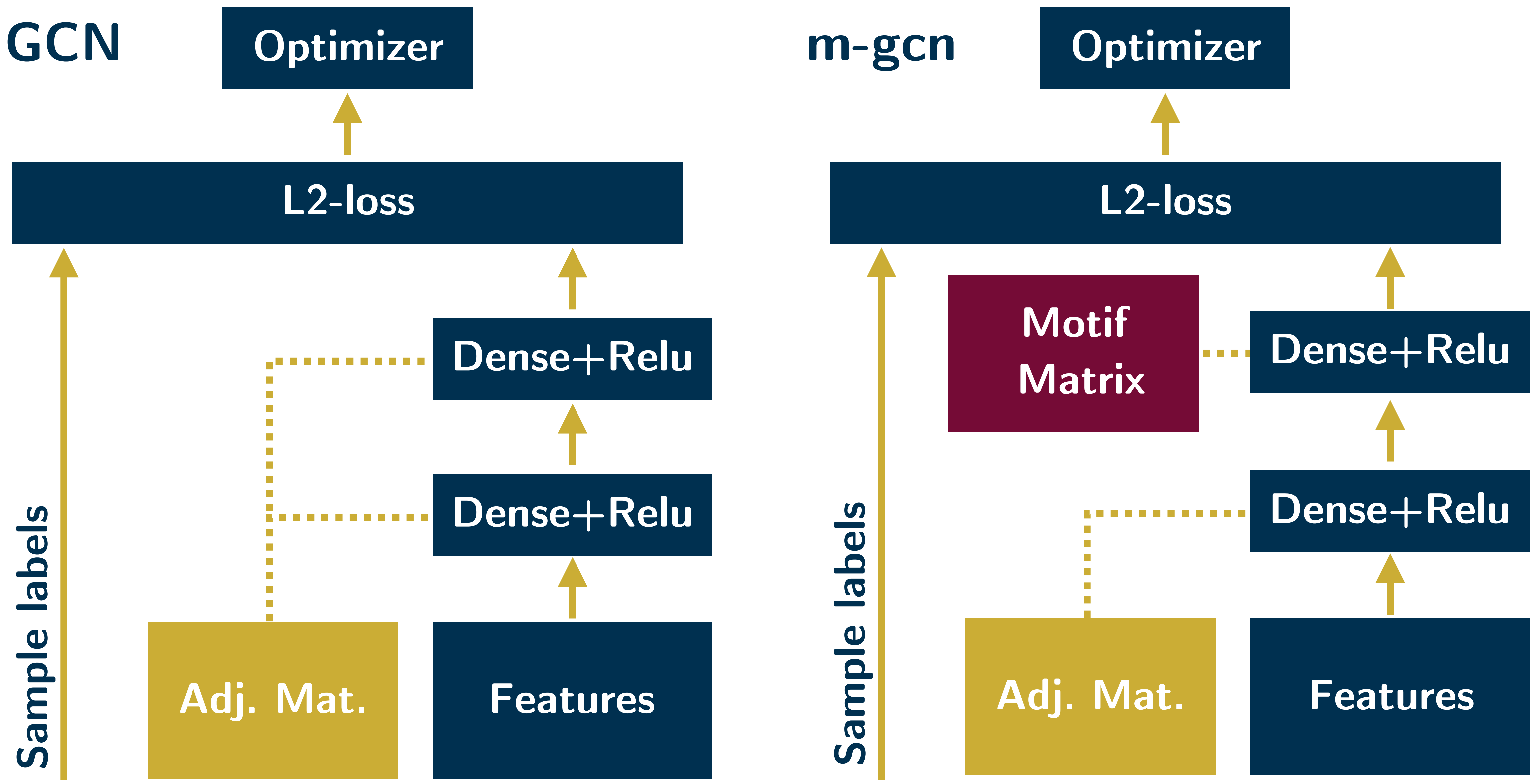
- Bottleneck at context generation.

Node multi-label classification

- Labels are interleaved.



Motif laplacian for graph convolutional networks



(Kipf, 2016) (Kipf, 2017)

Graph convolutional networks

- Define a wavelet basis with graph structure.
- Let the neural network learns the filter parameter for the defined basis.

Some filter Signal

$$g_\theta \star x = U g_\theta U^\top x$$

Eigenvectors matrix of the graph Laplacian

$$g_\theta \star x \approx \theta \left(I_N + D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \right) x$$

(Kipf, 2016)

Node label classification

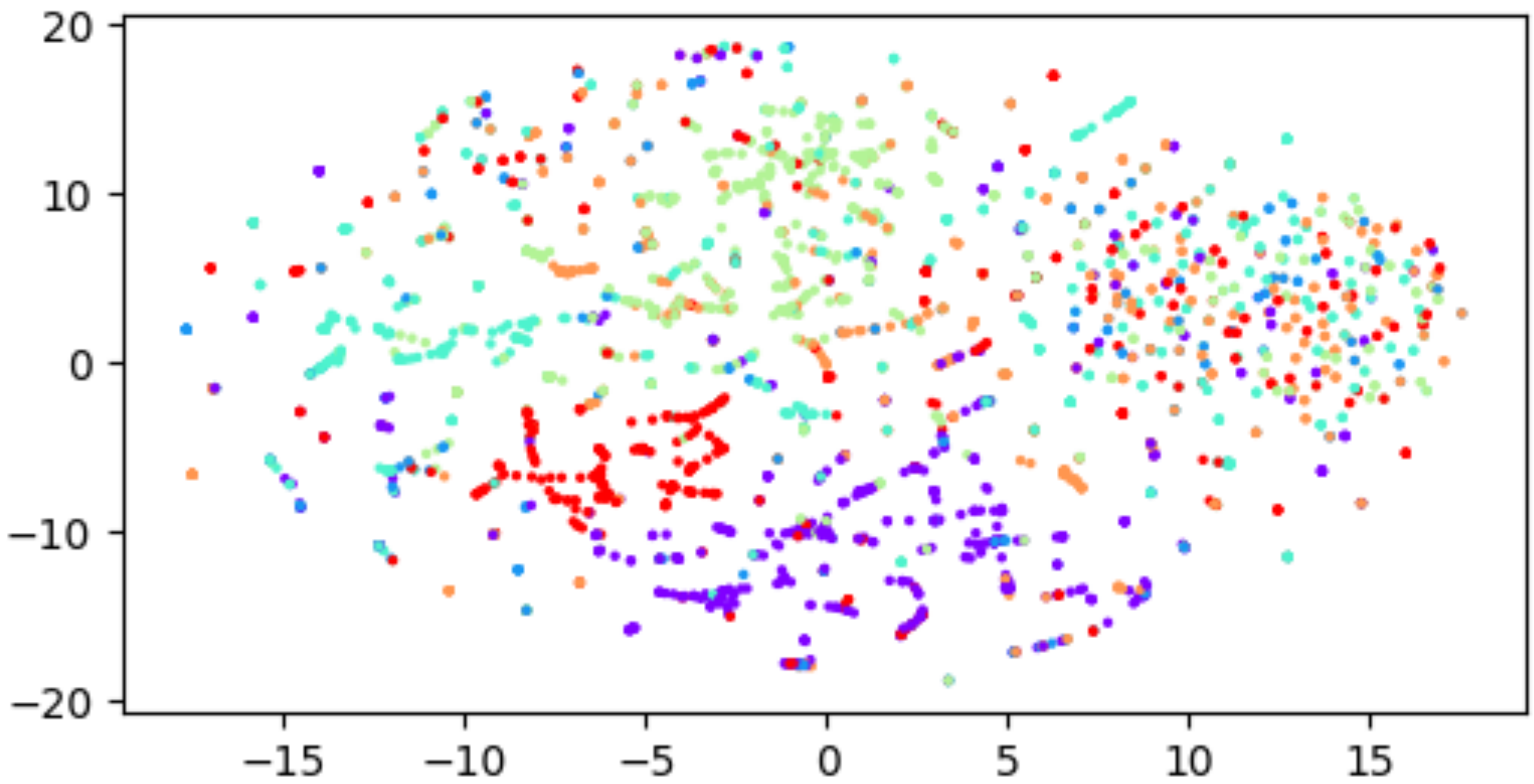
- Citation networks: Cora, citeseer, pubmed. - f1 and acc.
- 3 to 6 classes, each node has exactly one label.

Methods	Cora	Citeseer	Pubmed
Spectral Clus.	59.5	42.7	65.3
DeepWalk	67.2	43.2	65.3
GCN	81.5	70.3	79.0
m-gcn (ours)	**83.3**	**71.4**	**80.4**

- Only up to 2 percentage points improvement.
- Motif search and co-occurence matrix construction.

Node label classification

- Citeseer labels.



References

(Mikolov, 2013) Distributed Representations of Words and Phrases and their Compositionality.

(Bengio, 2013) Deep Learning of Representations: Looking Forward.

(Ng, 2001) On Spectral Clustering: Analysis and an algorithm.

(Perozzi, 2014) DeepWalk: Online Learning of Social Representations.

(Tang, 2015) LINE: Large-scale Information Network Embedding.

(Cao, 2015) GraRep: Learning Graph Representations with Global Structural Information.

(Grover, 2016) node2vec: Scalable Feature Learning for Networks

(Yang, 2016) Semi-Supervised Classification with Graph Convolutional Networks

(Alon, 2007) Network motifs: theory and experimental approaches.

(Benson, 2016) Higher-order organization of complex networks.

(Kipf, 2016) Variational Graph Auto-encoder.

(Kipf, 2017) Semi-Supervised Classification with Graph Convolutional Networks