GRAPH EMBEDDINGS AND DATA MODELING: A PRIMER

By: Samira Maghool

Postdoc researcher at University of Milan,

Department of Computer Science

DATABASES AT A GLANCE

Relational databases: working with tables and join clauses as a standard solution to organize well-structured, a typically stable, data

NoSQL databases: modern alternatives for data that doesn't fit the relational paradigm:

1) Key-value databases: simple, dictionary-style lookups for basic storage and retrieval,
2) Document databases: Storing all of an item's data in flexible, self-describing structures,
3) **Graph databases**: mapping relationships by focusing on how connections between data are meaningful,
4) Time series databases: tracking value changes over time

IS THERE A UNIQUE SOLUTION?

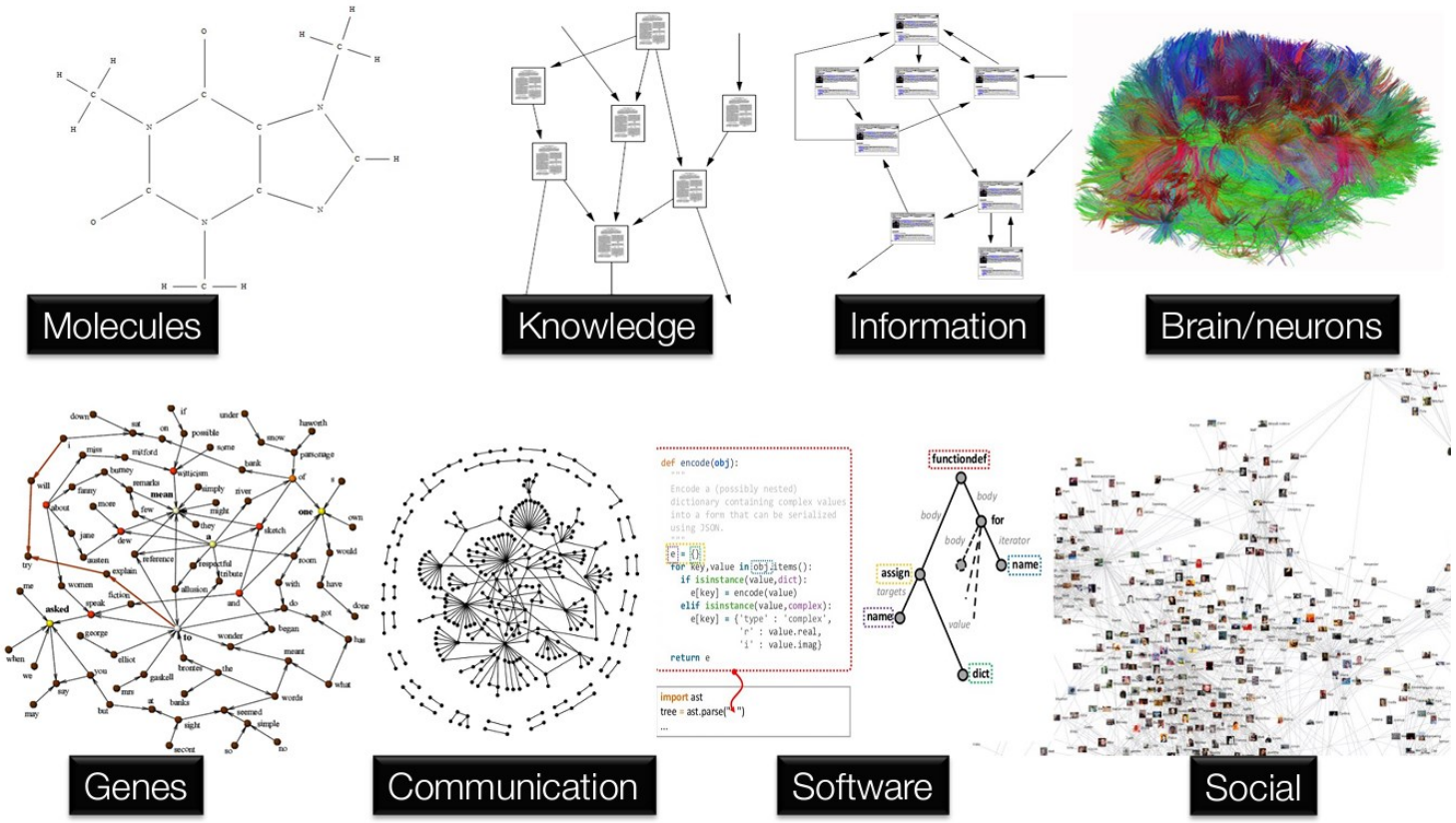
What if we have missed some of information, such as entities, relations?

What will happen to the learning/prediction tasks?

What if we have multiple type of data?

GRAPHS CONNECT THINGS

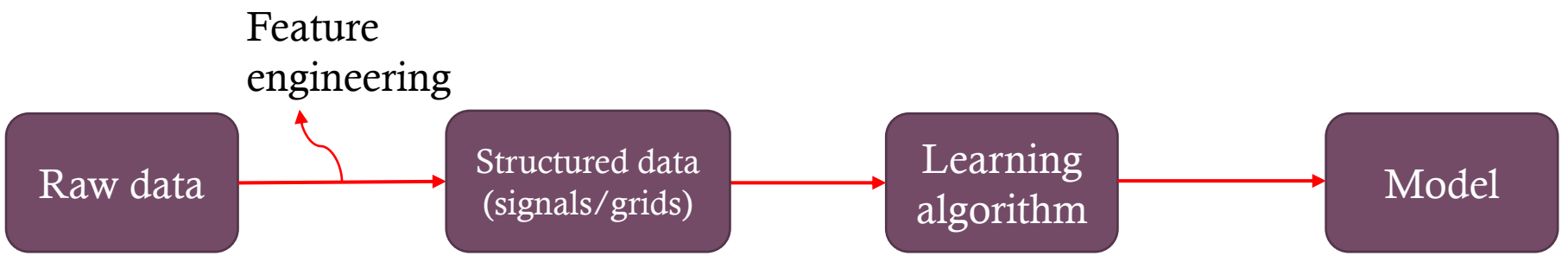
Considering the graph databases as a consistent solution to study noSQL databases could be inspiring in leveraging the graph representation of information around us…



HOW COULD WE PROCEED WITH REPRESENTATION LEARNING USING GRAPHS?

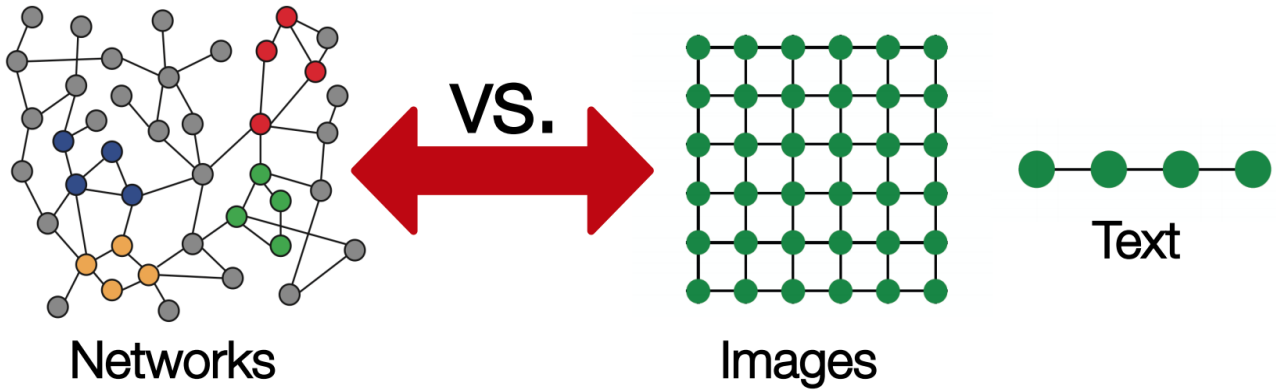
There might be a solution in using ML/DL !!!

- ML Toolbox (classical neural networks) is typically designed for simple sequences and grids (feature vectors)

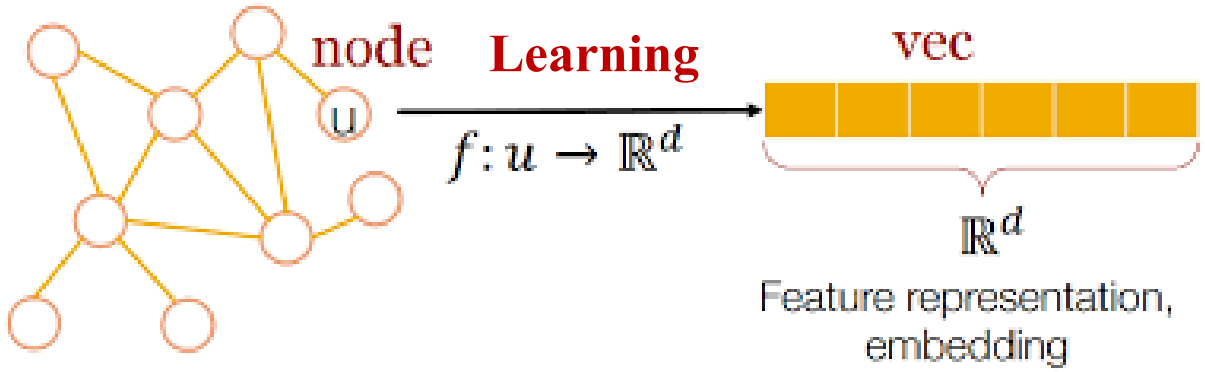


Networks are complex:

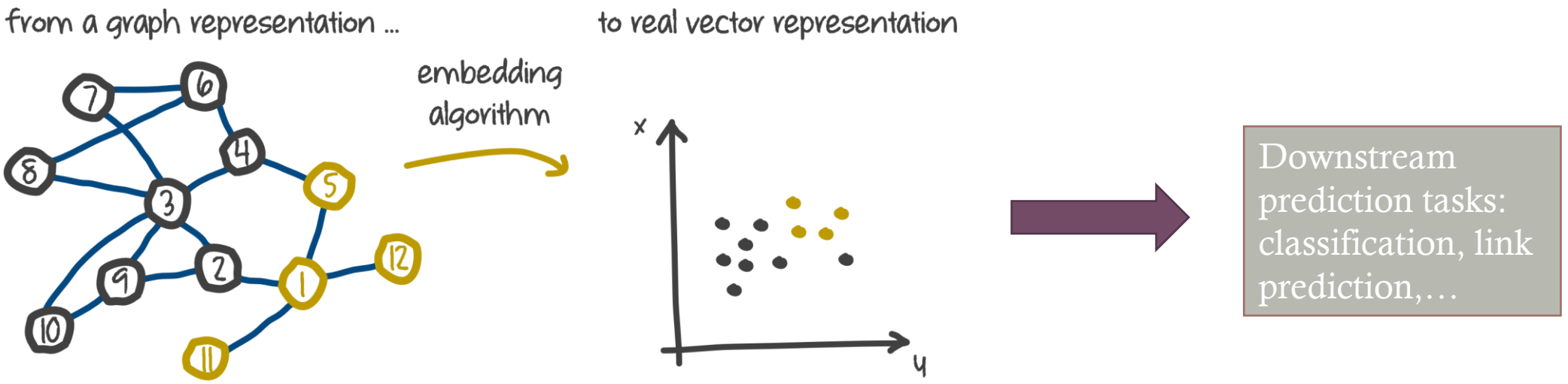
1. Arbitrary size and complex topological structure (i.e., no spatial locality like grids)
2. No fixed node ordering or reference point
3. Often dynamic and have multimodal features



WHAT IS REPRESENTATION LEARNING FOR GRAPHS?



Graph embedding helps in transforming nodes, edges, and their features into a vector space while partially preserving properties of their original graph structure.



The general goal is that connected nodes in the graph kept closer in the latent space

https://gearons.org/posts/mage

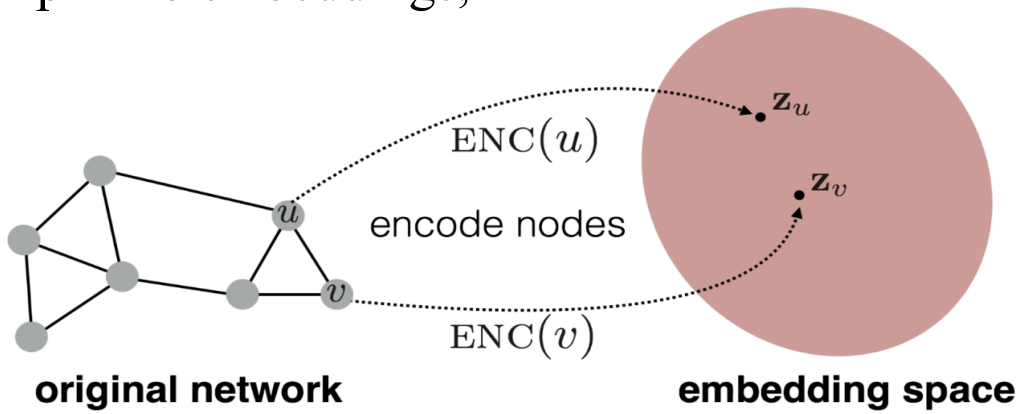
Three **common steps** are needed in defining an embedding procedure:

❖ Define an encoding procedure mapping the nodes of a graph into embeddings,

$$\mathrm{ENC}(v) = \mathbf{z}_v$$

d-dimensional embedding

node in the input graph



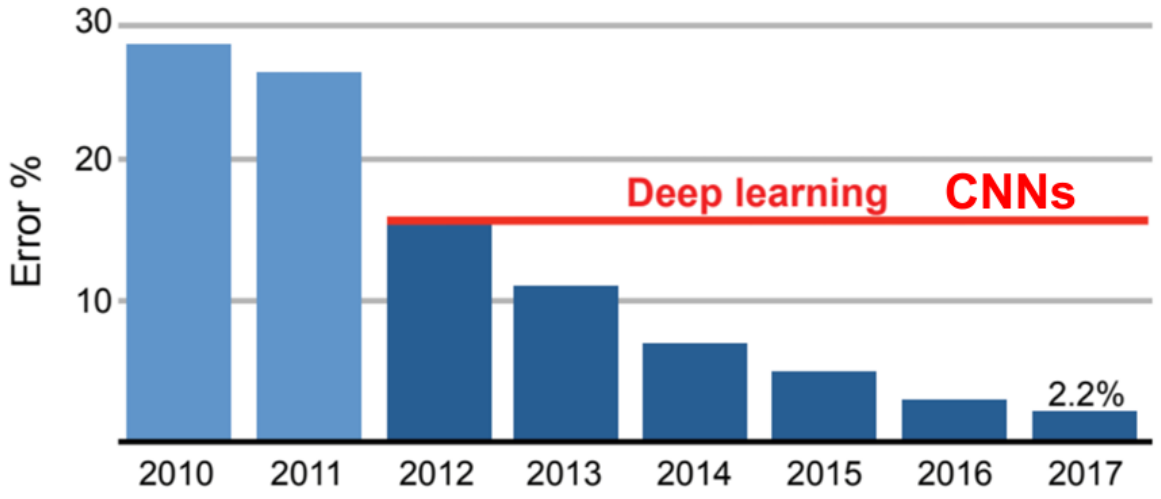
❖ Define a node similarity function(i.e., a measure of similarity in the original network)

$$similarity(u, v) \approx z_u^T z_v$$

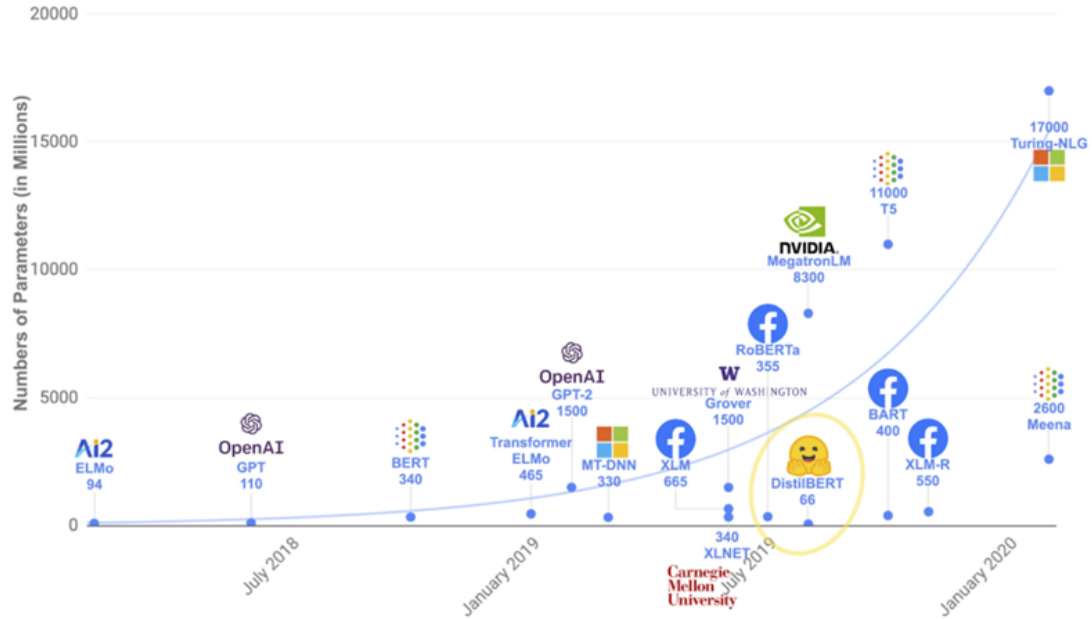
❖ Optimize the parameters of the encoder so that maximizes the similarity function.

WE ARE AT THE AI RENAISSANCE

- The deep learning revolution caused breakthroughs in image recognition by Convolutional Neural Networks and in natural language understanding fueled by Transformers, …



Jure Leskovec (@jure), Stanford University

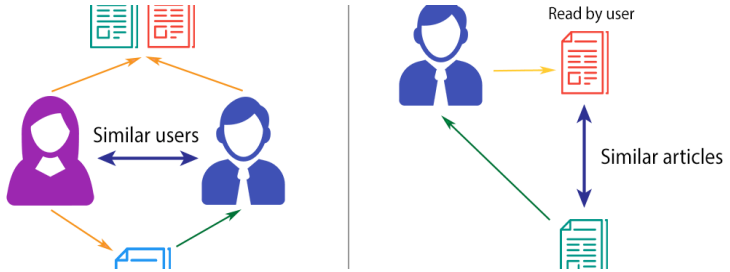


Jure Leskovec (@jure), Stanford University

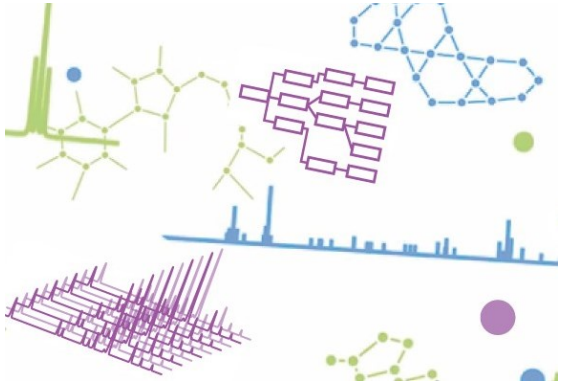
WE NEED TO DEVELOP NEURAL NETWORKS THAT ARE MUCH MORE BROADLY APPLICABLE

In other words, Graphs are the new frontier of deep learning

APPLICATION OF DL ON NETWORKS



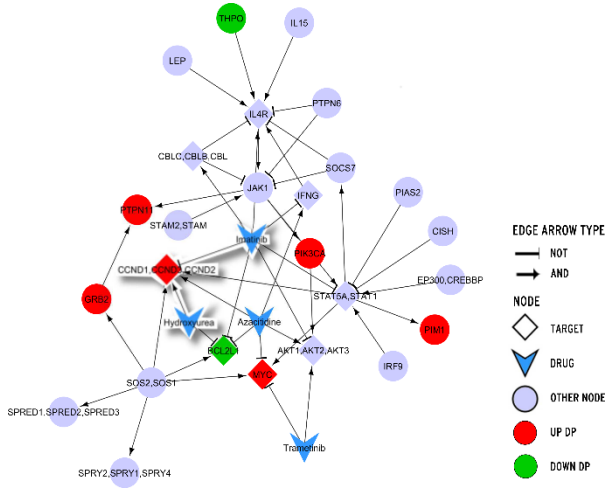
Recommender systems



Chemistry

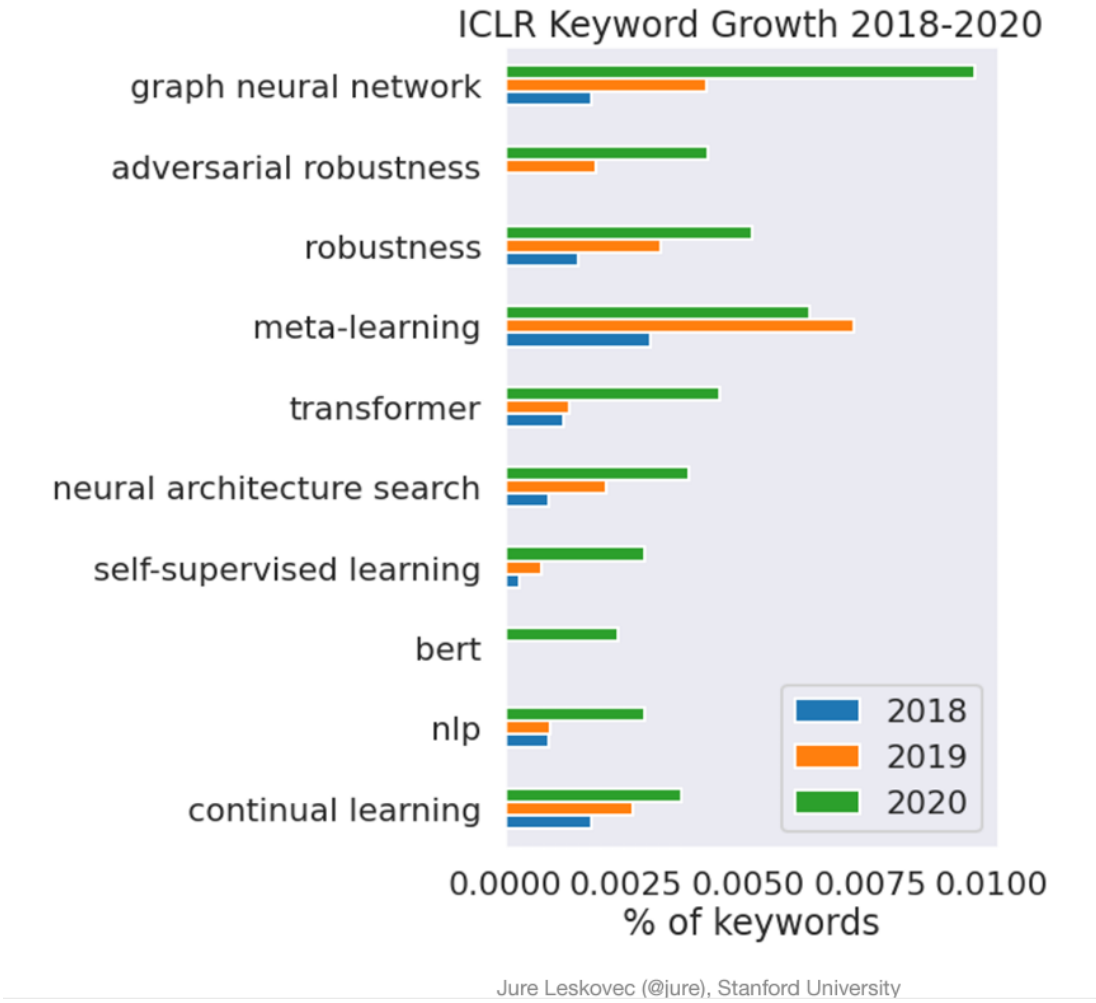
Fraud detection

Drug repurposing

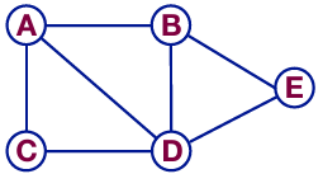


Augmented reality

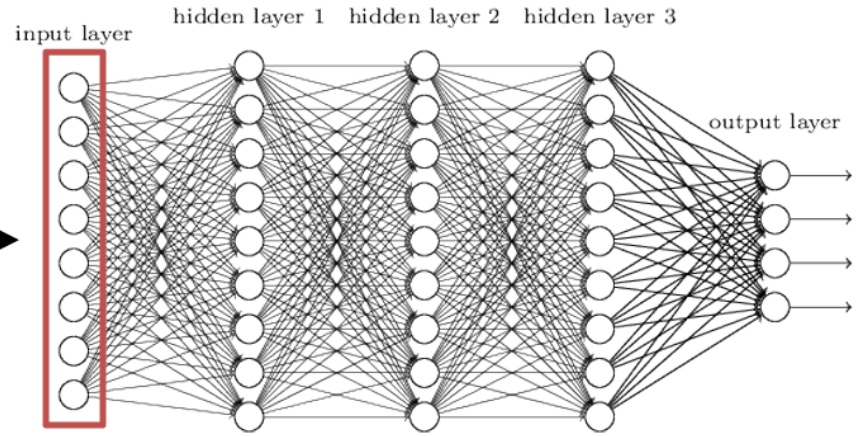
HOT TOPICS:


ICLR Keyword Growth 2018-2020
graph neural network
adversarial robustness
robustness
meta-learning
transformer
neural architecture search
self-supervised learning
bert
nlp
continual learning
2018
2019
2020
0.0000 0.0025 0.0050 0.0075 0.0100
% of keywords
Jure Leskovec (@jure), Stanford University

A NAÏVE APPROACH FOR LEARNING:

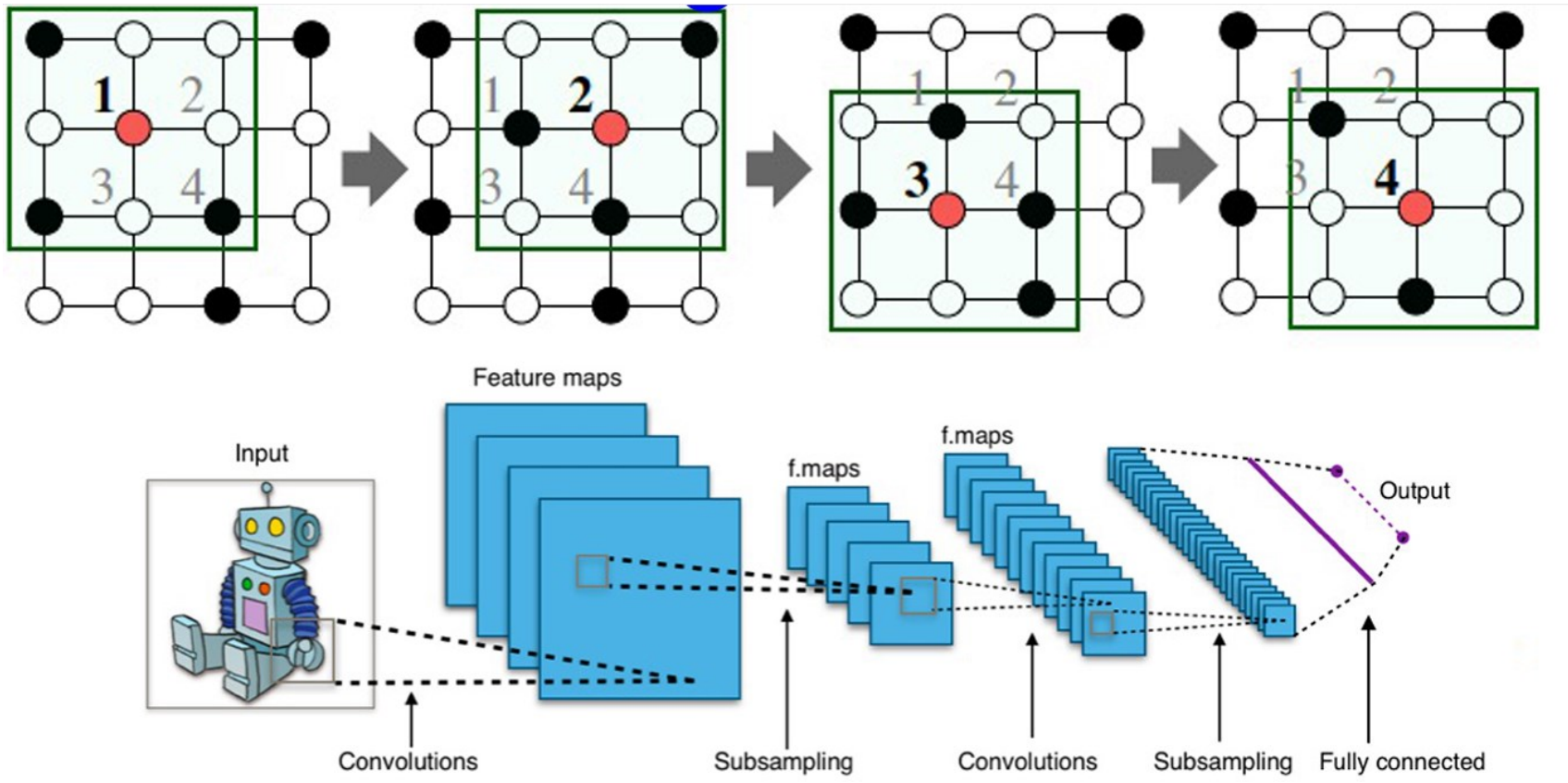


	A	B	C	D	E	Feat	
A	0	1	1	1	0	1	0
B	1	0	0	1	1	0	0
C	1	0	0	1	0	0	1
D	1	1	1	0	1	1	1
E	0	1	0	1	0	1	0

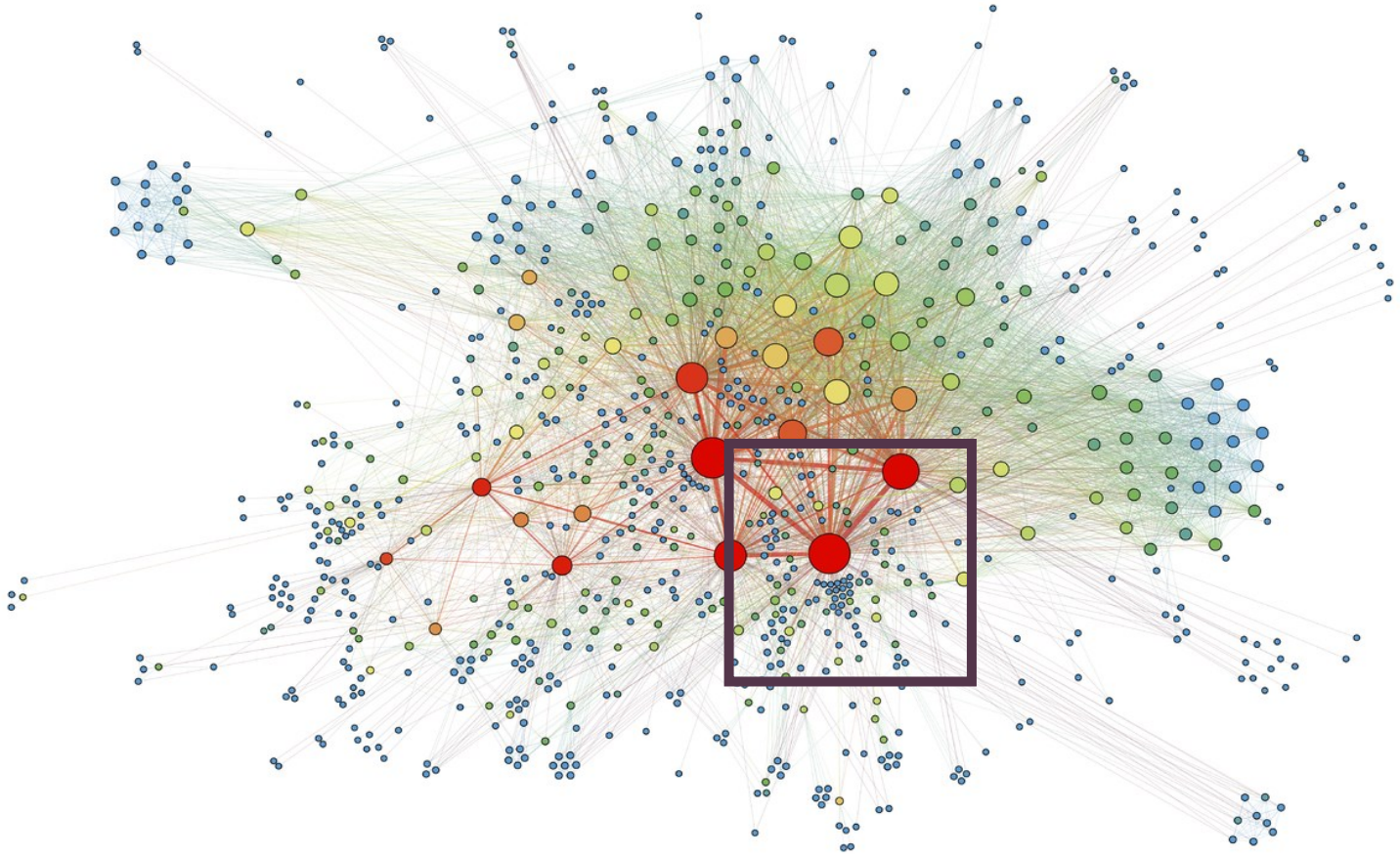


?

THE IDEA IS USING CNN…

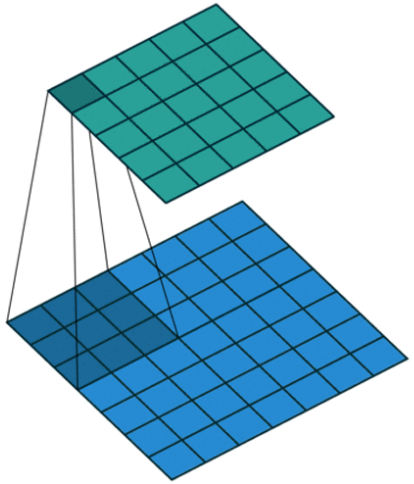


HOW IS POSSIBLE TO USE CNN ON GRAPHS?

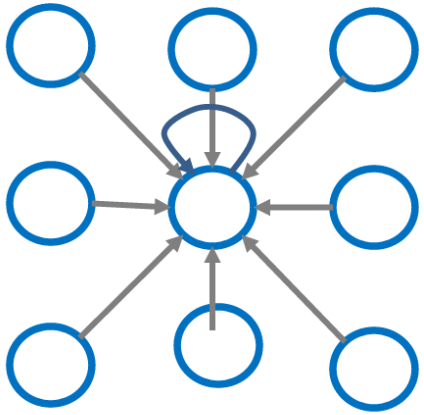


FROM IMAGES TO GRAPHS:

(CNN) layer with 3x3 filter:



Image

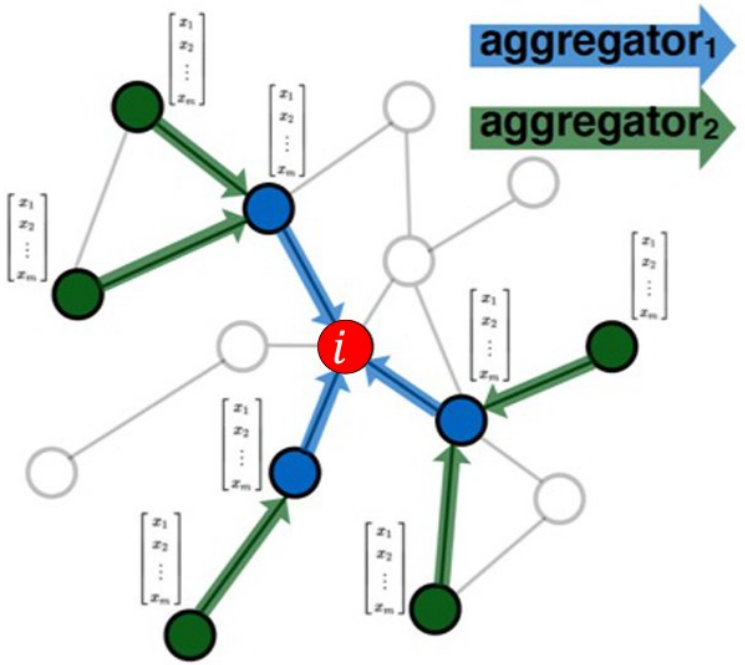


Graph

NETWORK IS A COMPUTATION GRAPH AND LEARNS HOW TO PROPAGATE INFORMATION

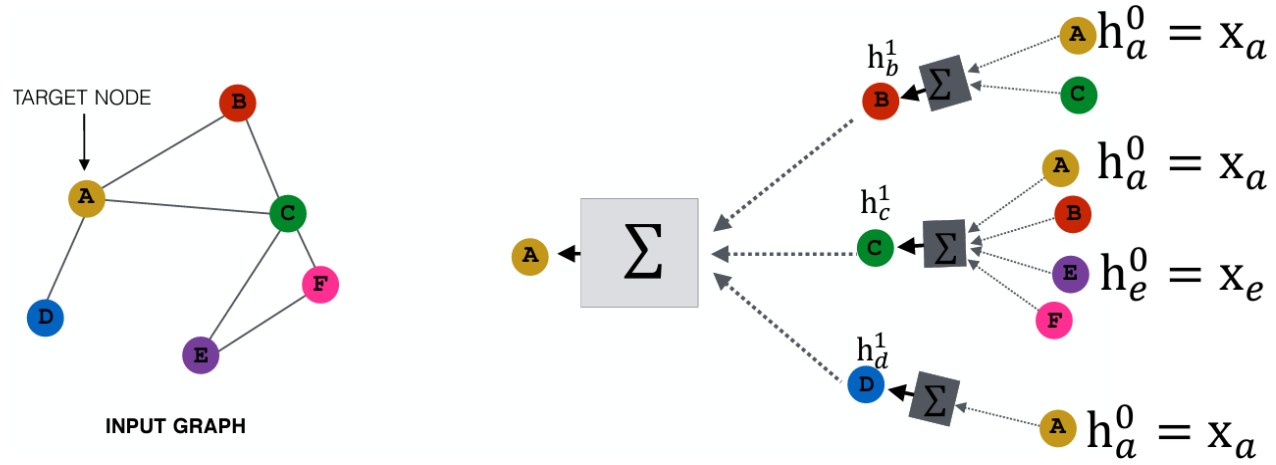
Transform "messages" h_i from neighbors : $W_i h_i$

Add them up: $\sum W_i h_i$



GRAPH NEURAL NETWORKS

- Each node is a computation unit
- Each edge in this graph is a transformation/aggregation function

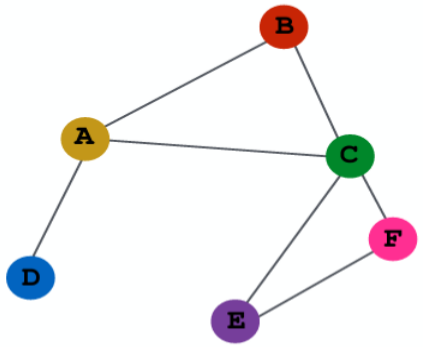


Neighborhood aggregation

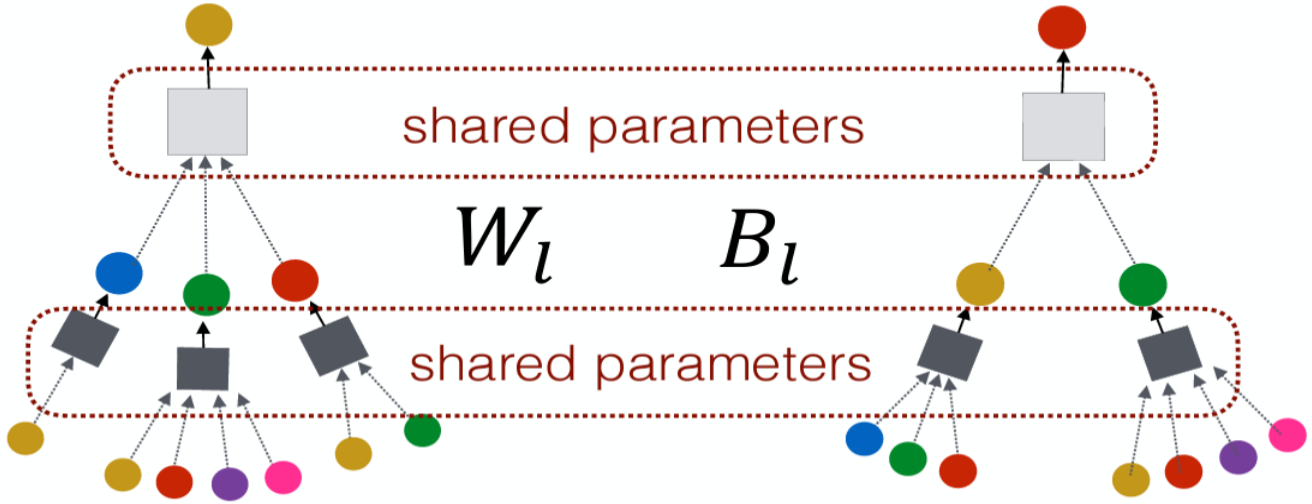
$$h_v^{(l+1)} = \sigma(W_l \sum_{u \in N(v)} \frac{h_u^{(l)}}{|N(v)|} + B_l h_v^{(l)})$$

Scarselli et al. 2005. The Graph Neural Network Model . IEEE Transactions on Neural Networks

INDUCTIVE…

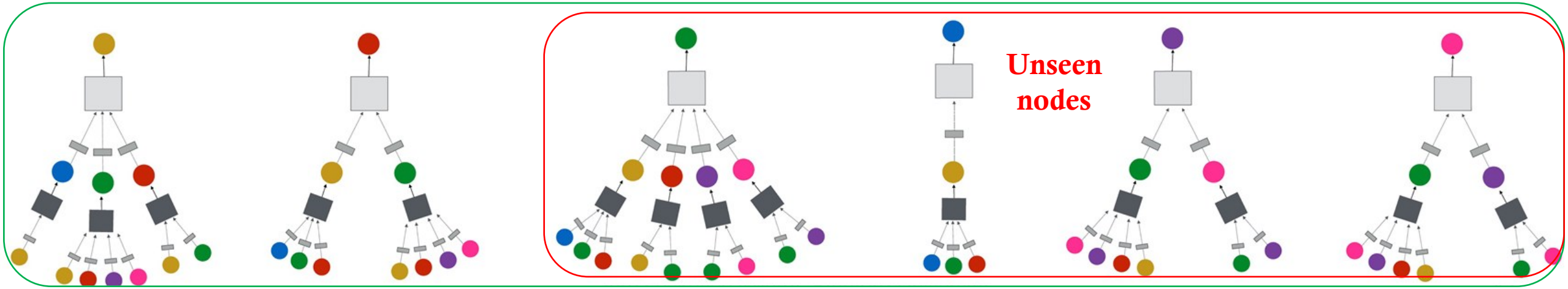


INPUT GRAPH



Compute graph for node A

Compute graph for node B

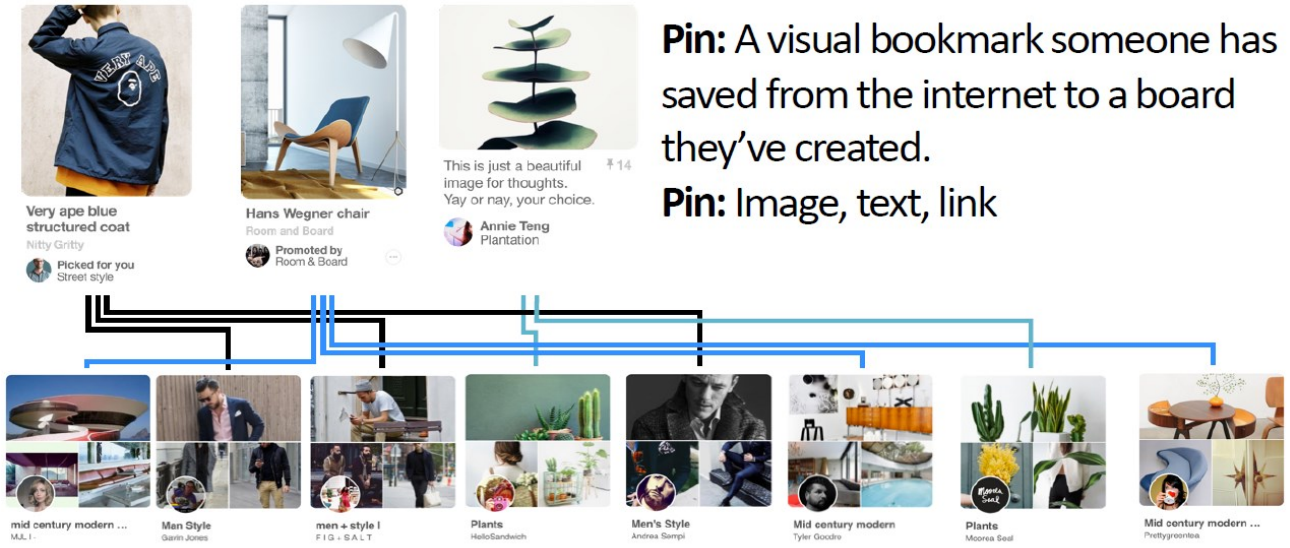


CAPABILITIES…

- Inductive: Generate embeddings for nodes as needed. Even for nodes we never trained on!
- No manual feature engineering is needed
- End-to-end learning results in optimal features
- Any graph machine learning task: Node-level, link-level, entire graph-level prediction
- Scalable to billion node graphs!
- Other Deep Learning architectures assume fixed input (matrix, sequence), GNN doesn't.

DATA PRUNING AND GRAPH REFORMATION

- Generally speaking, the preferred GNN approach, is highly dependent on the favored learning task, GraphSAGE works well with network of pictures (Pinterest)
- static/dynamic network structure
- Obtaining high accuracy may require graph reformation considering a coarse-grained picture of graph, pre-defined pattern detection.



CONCLUSION

- Graph connect things
- Graph databases are very powerful tools
- Probabilistic knowledge extraction method is needed for completing missed information or prediction task
- Generic knowledge queries are difficult, for example the similarity of large group of nodes
- Graph embedding could bring up a unified tool for multimodal data analysis

THANKS FOR YOUR ATTENTION