Graph Convolutional Networks

Mì AI

What's up?

+ What is a Graph?

+ NetworkX package in Python

+ Graph Neural Network (GNN)

+ Graph Convolution Network (GCN)
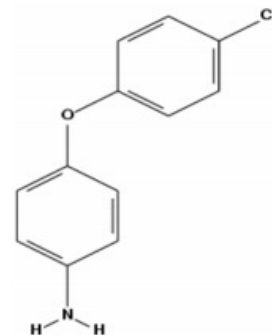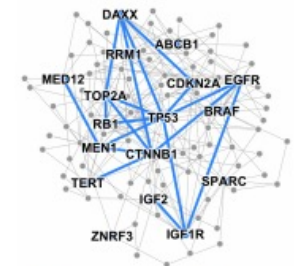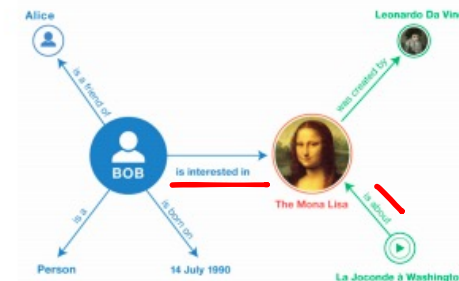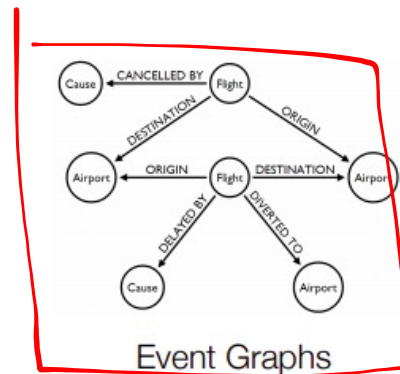
+ Handson

What's Graph?

A Graph is the type of data structure that contains nodes and edges. A node can be a person, place, or thing, and the edges define the relationship between nodes. The edges can be directed and undirected based on directional dependencies.
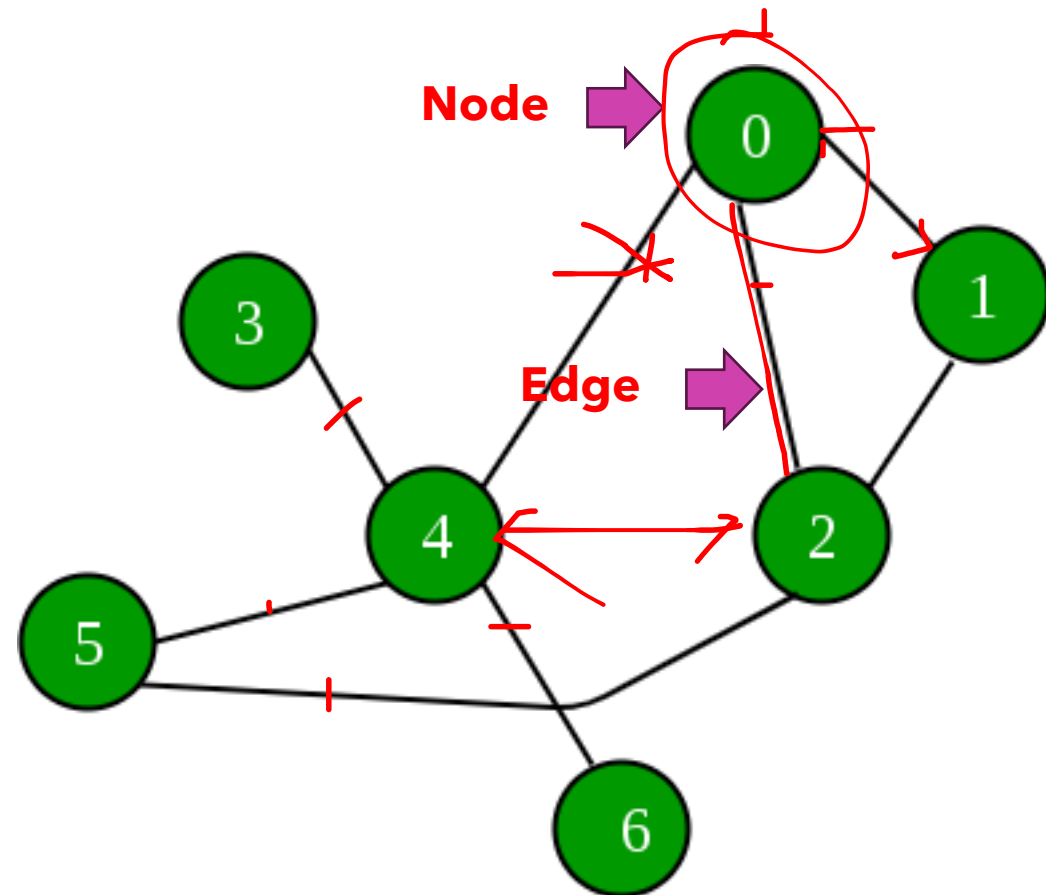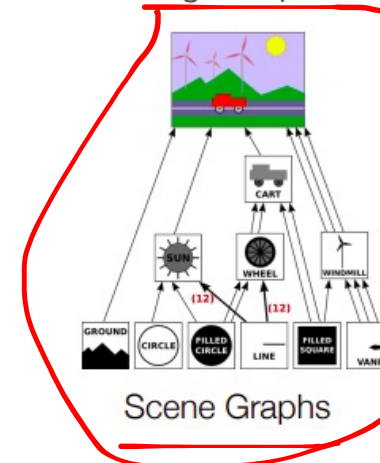
What's Graph?

+ <u>Nodes can be</u>:
+ <u>Edges can be</u>:

Event Graphs

Knowledge Graphs
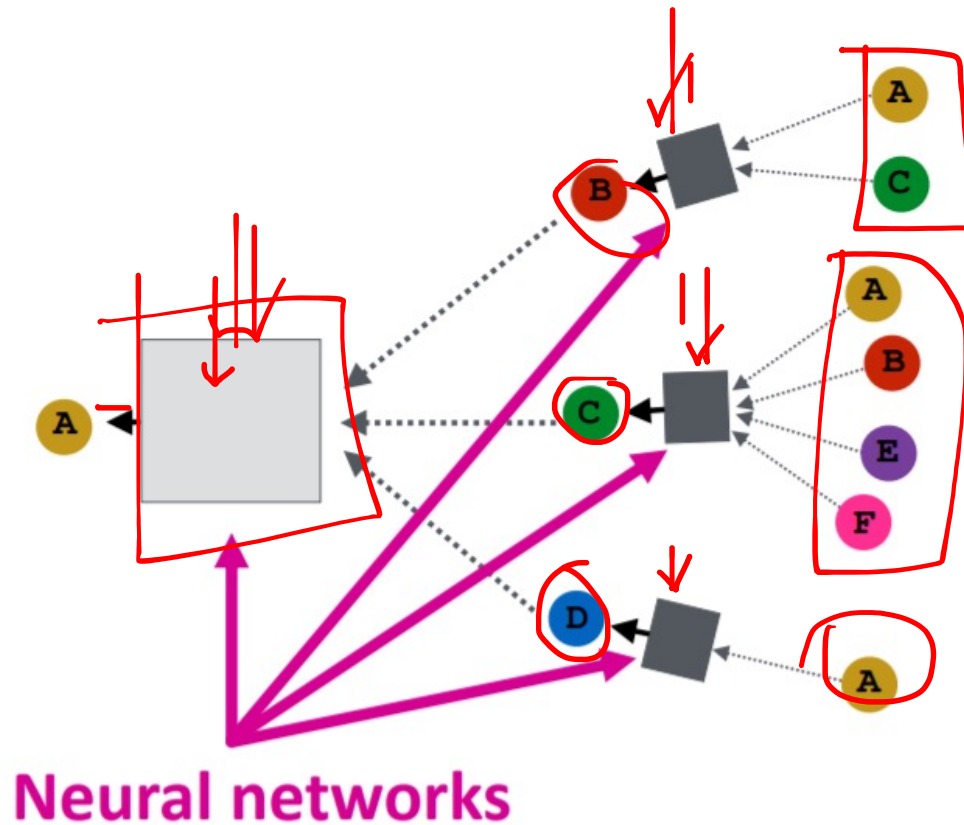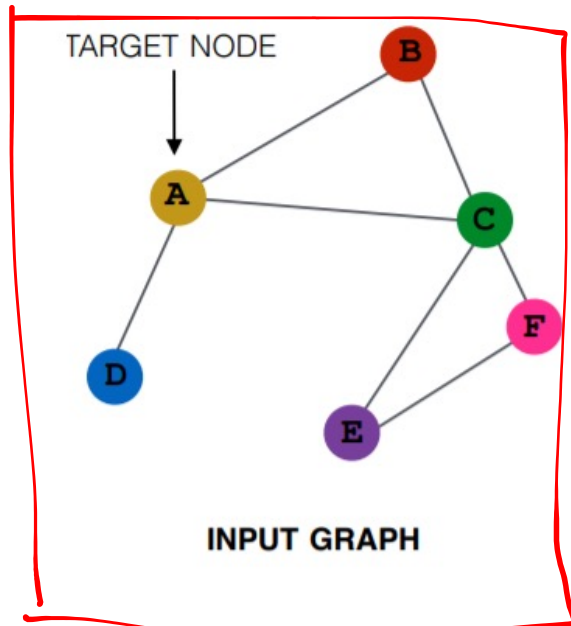

Disease pathways

Molecules

Scene Graphs

Cell-cell similarity networks

NetworkX

NetworkX is a Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks.

https://networkx.org/

GNN

Graph Neural Networks (GNNs) is a type of deep learning approach that performs inference on graph-described data.

GNN

The main goal of GNN is for each of the nodes to learn an embedding containing information about its surroundings. This process of embedding can be used for many applications like node labeling, node prediction, edge prediction, etc.
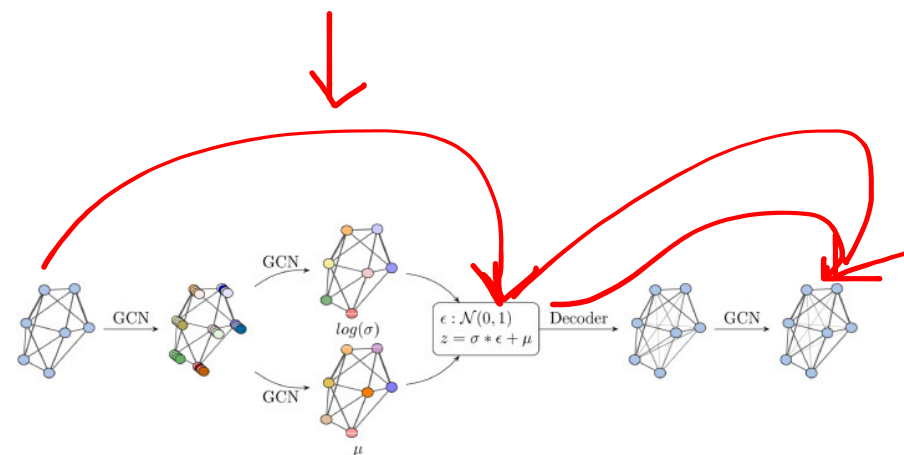
Thus, once we've assigned embeddings to each node, we may transform edges by adding feed-forward neural network layers and merge graphs with neural networks.

node

$f : u \rightarrow \mathbb{R}^d$

vec

$\mathbb{R}^d$

Feature representation, embedding

GNN
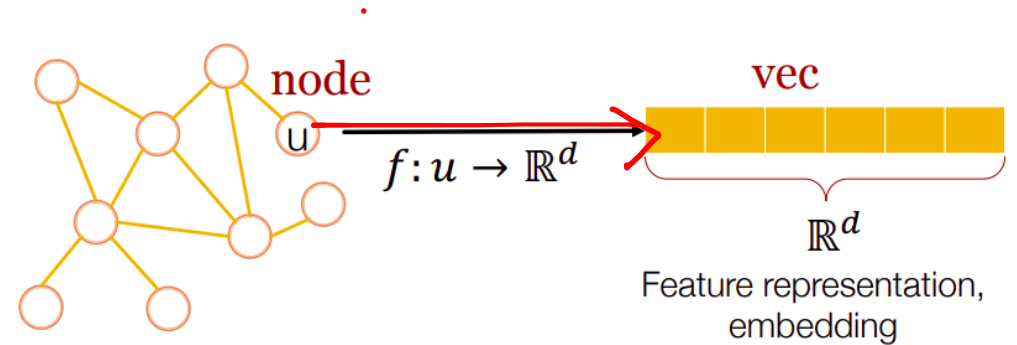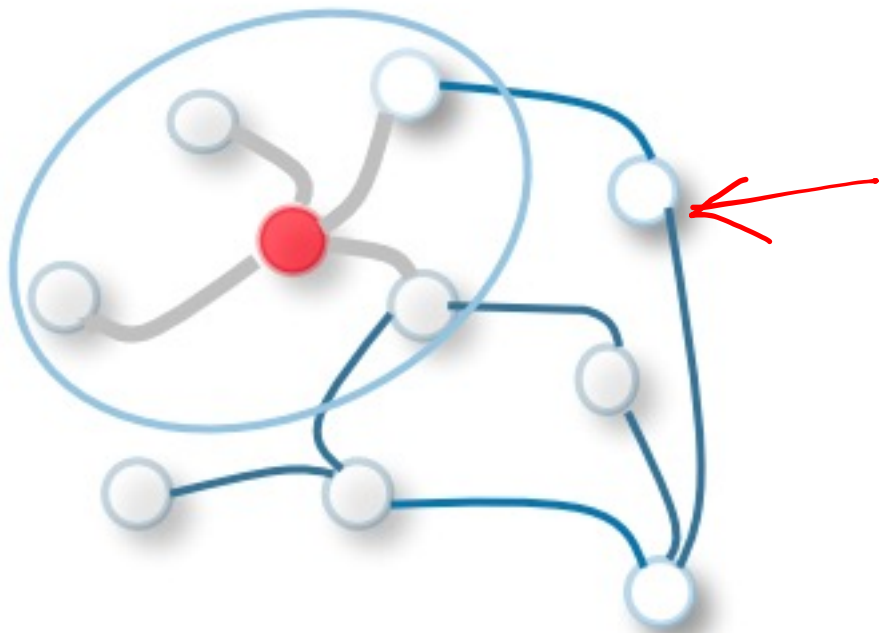
Graph Convolutional Network

Graph Auto-Encoder Network

Applications of GNN

- Graph Classification
- Node Classification
- Link Prediction
- Graph Clustering: Clustering node or clustering graphs
- Graph visualization

Applications of GNN

Applications of GNN

"To-buy-together" Recommendations

Query Product

Head SpeedX
(Djokovic Racket)

List 1 — Bad. Tennis rackets are substitutes instead of complements.

List 2 — Mediocre. Tennis balls are related complements but the results lack diversity.

List 3 — Good. Recommendations are well related and diverse.
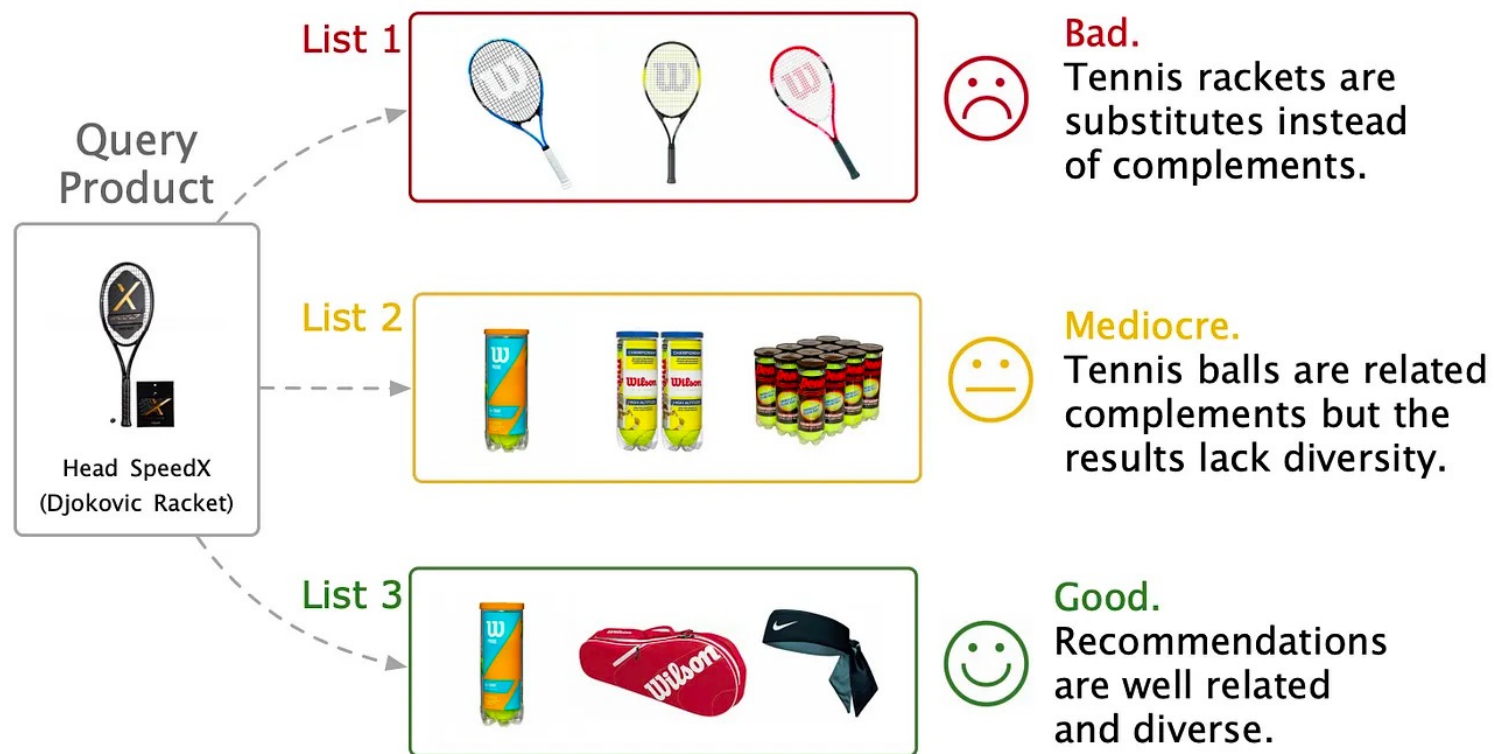
Applications of GNN

Human-object interaction

(i) Image   (ii) HOI candidates   (iii) HOI result

Joint Inference

(iv) Initial HOI graph   (v) Parse graph learning   (vi) Message passing   (vii) Final parse graph

GCN

GCN

TARGET NODE

INPUT GRAPH

Neural networks
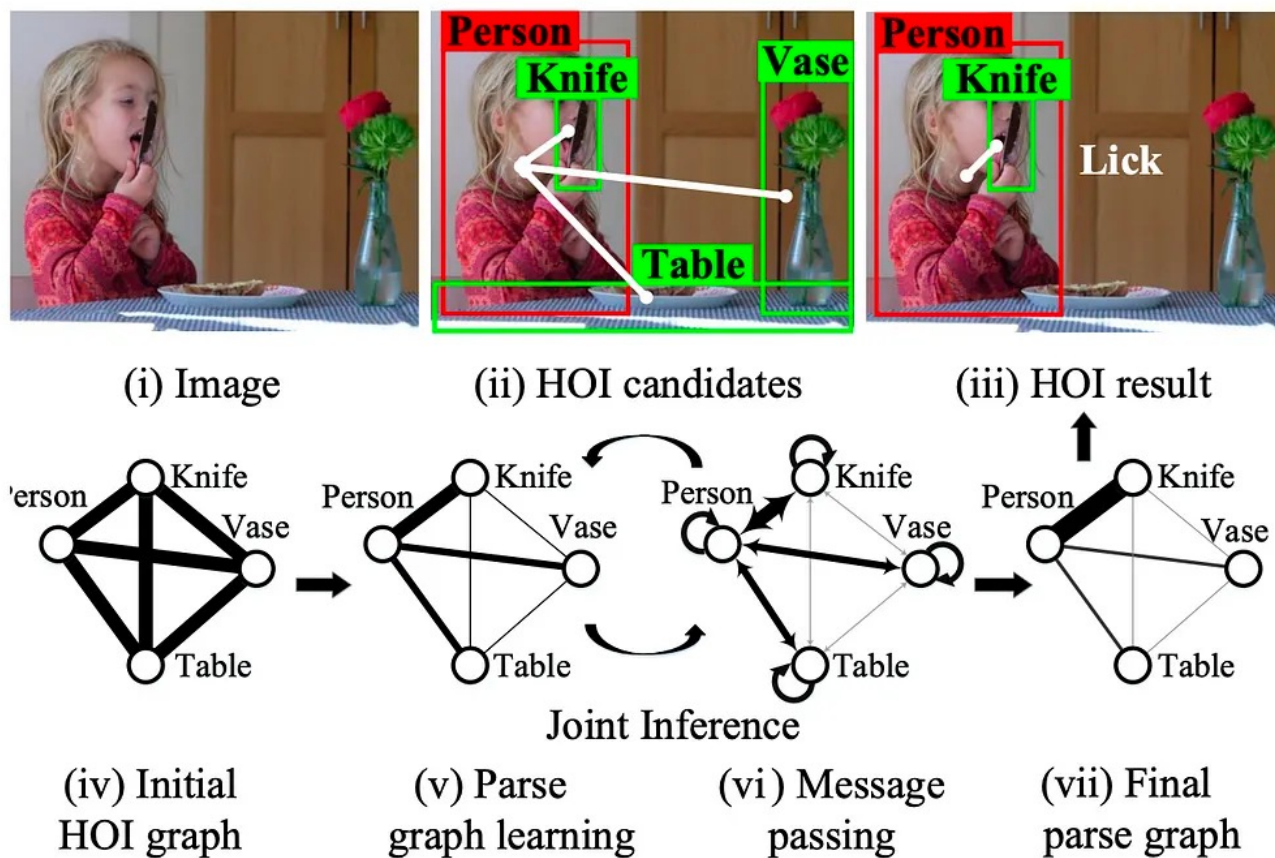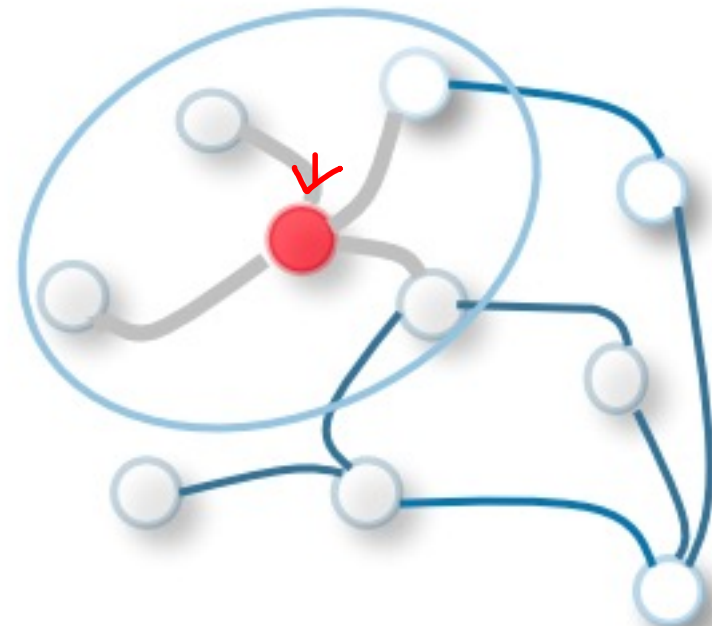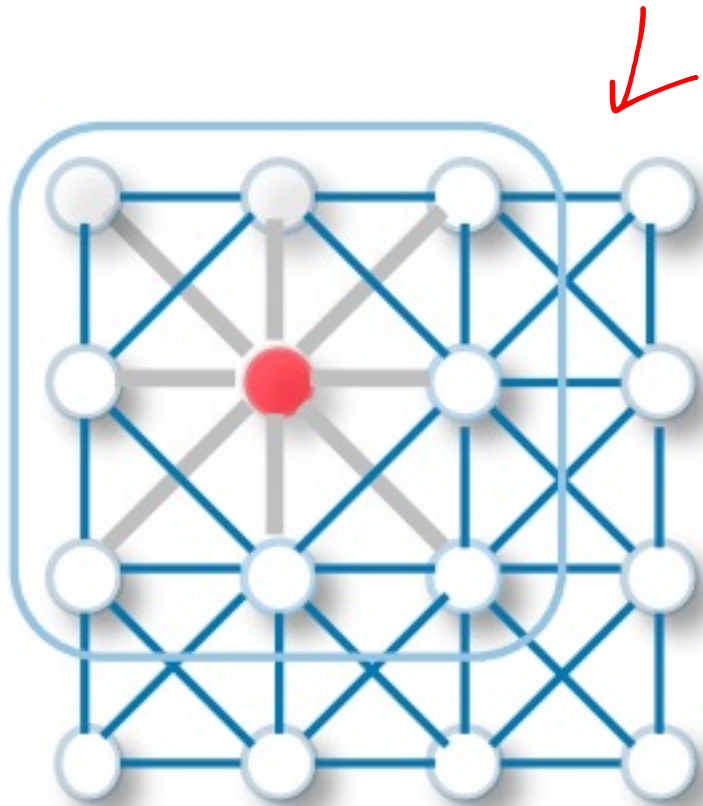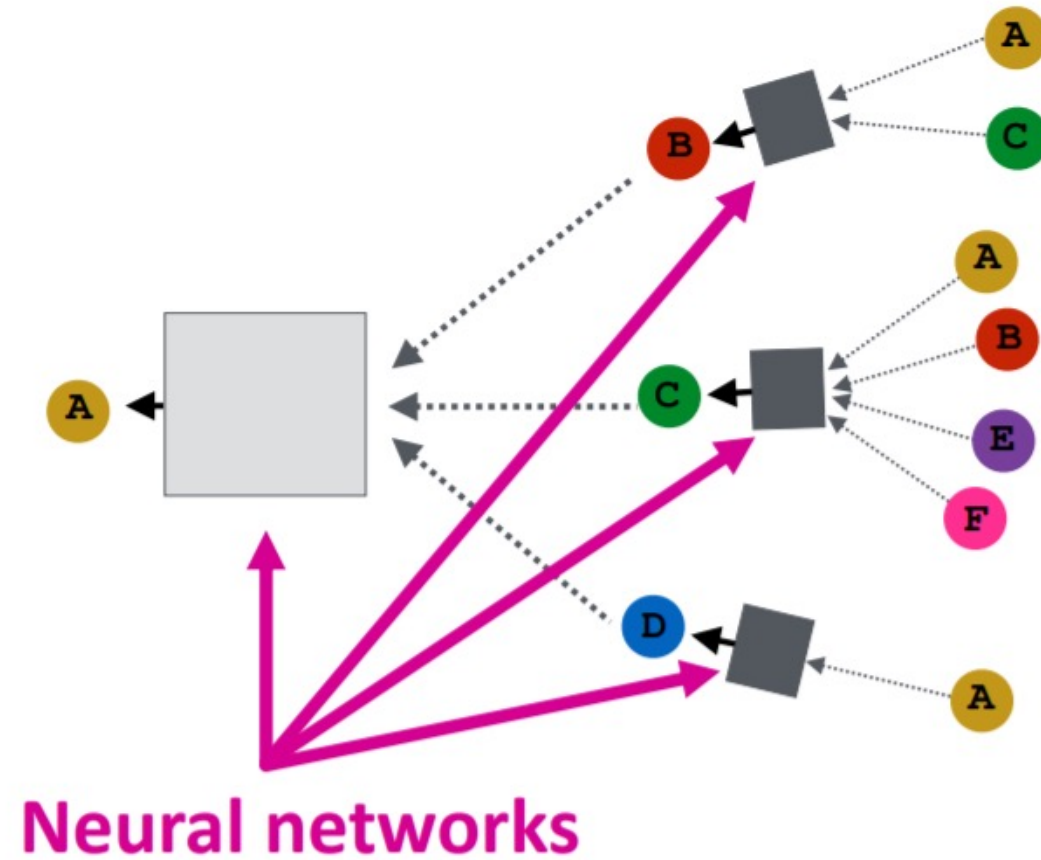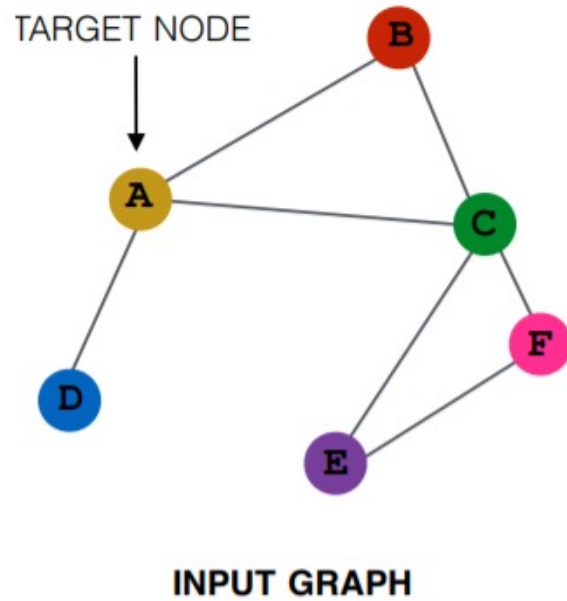
GCN

- **Intuition:** Network neighborhood defines a computation graph

Every node defines a unique computation graph!

INPUT GRAPH

GCN

GCN

TARGET NODE

what's in the box!?

INPUT GRAPH

GCN

GCN

INPUT GRAPH

$\mathbf{z}_A$

Need to define a loss function on the embeddings, $\mathcal{L}(z_u)$!
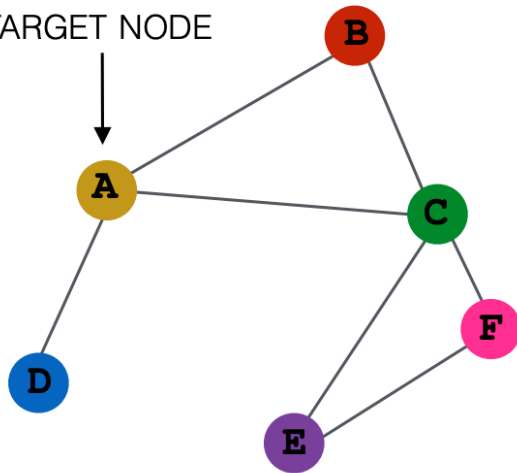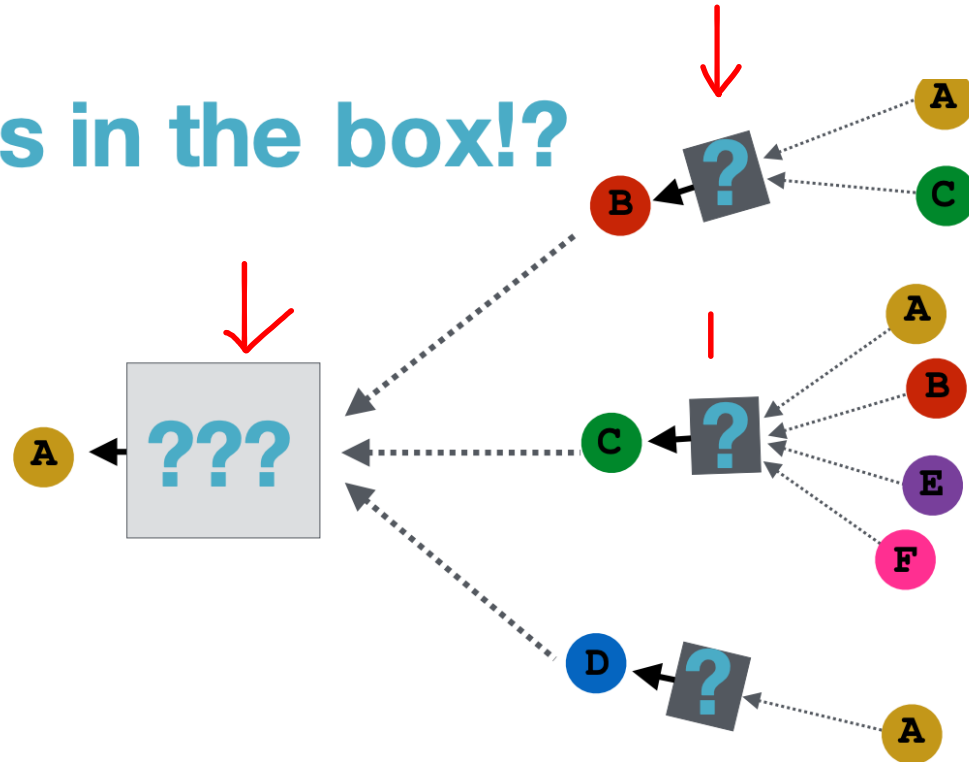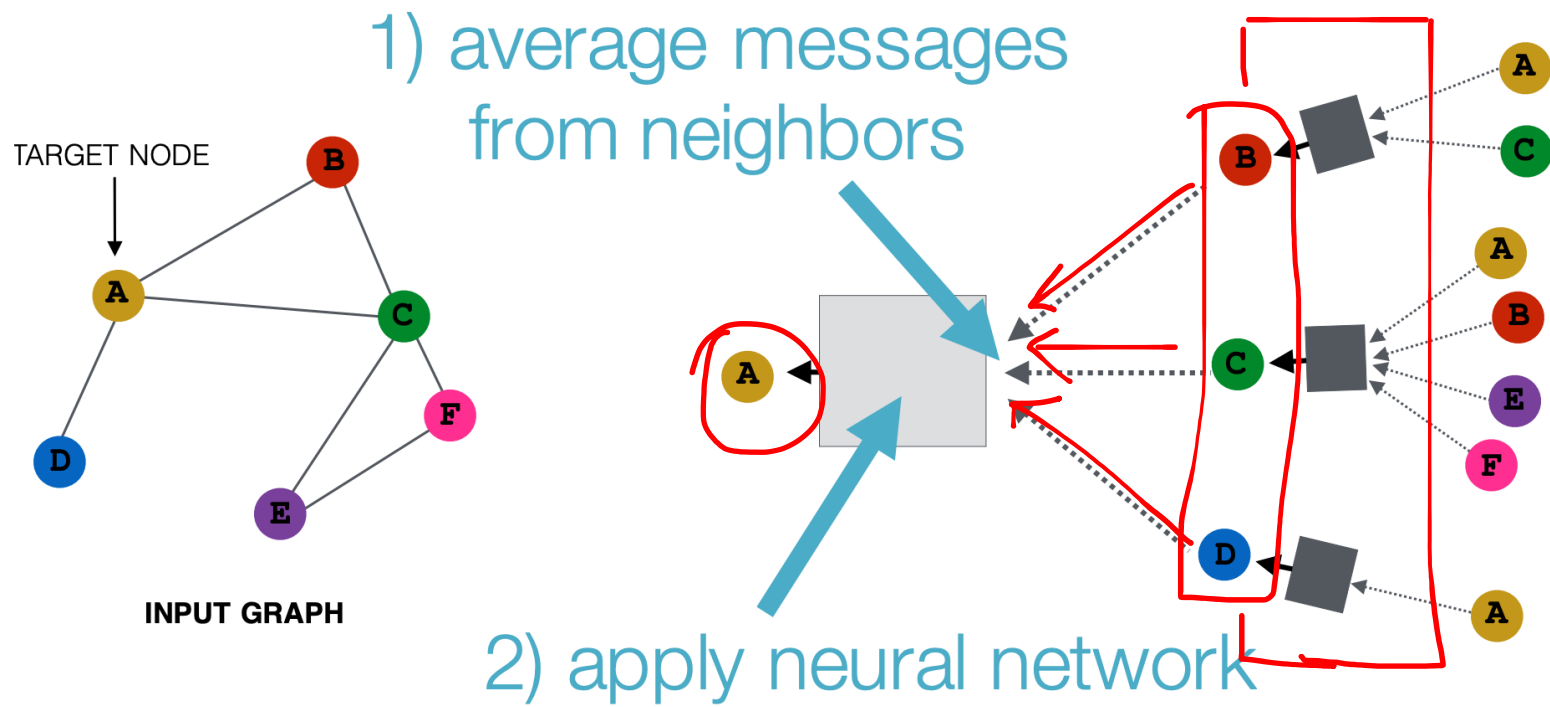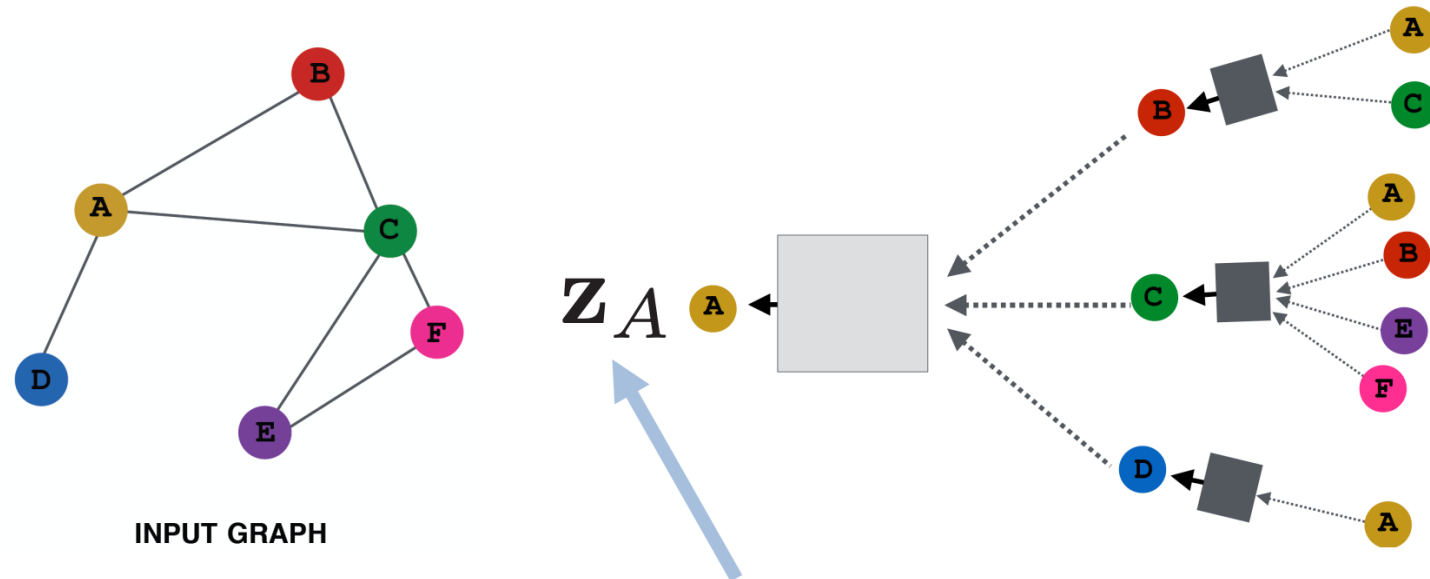
GCN
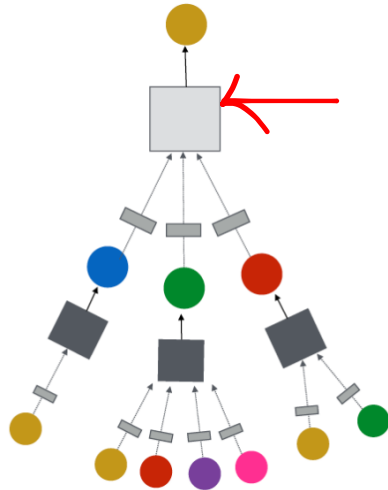
Human or bot?

Human or bot?
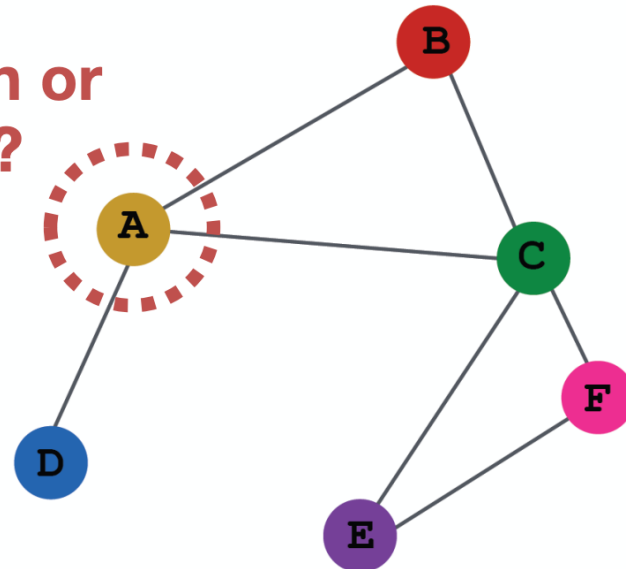
e.g., an online social network

GCN

AN INFORMATION EXTRACTION PIPELINE
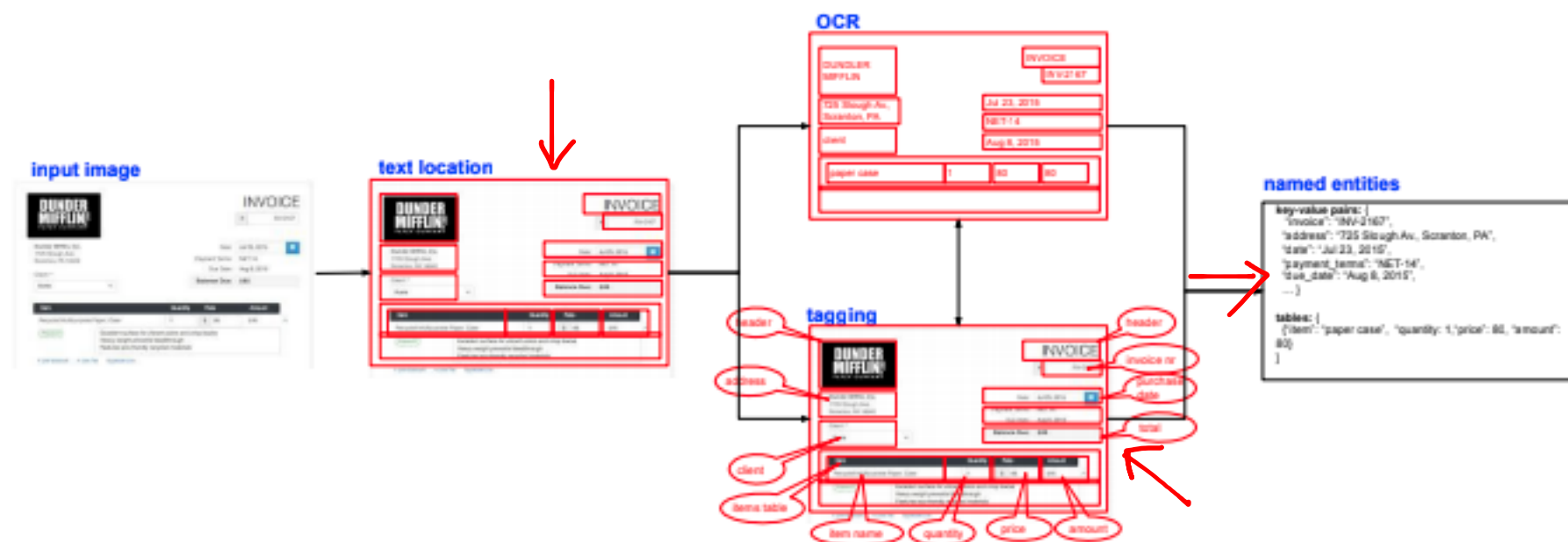
GCN
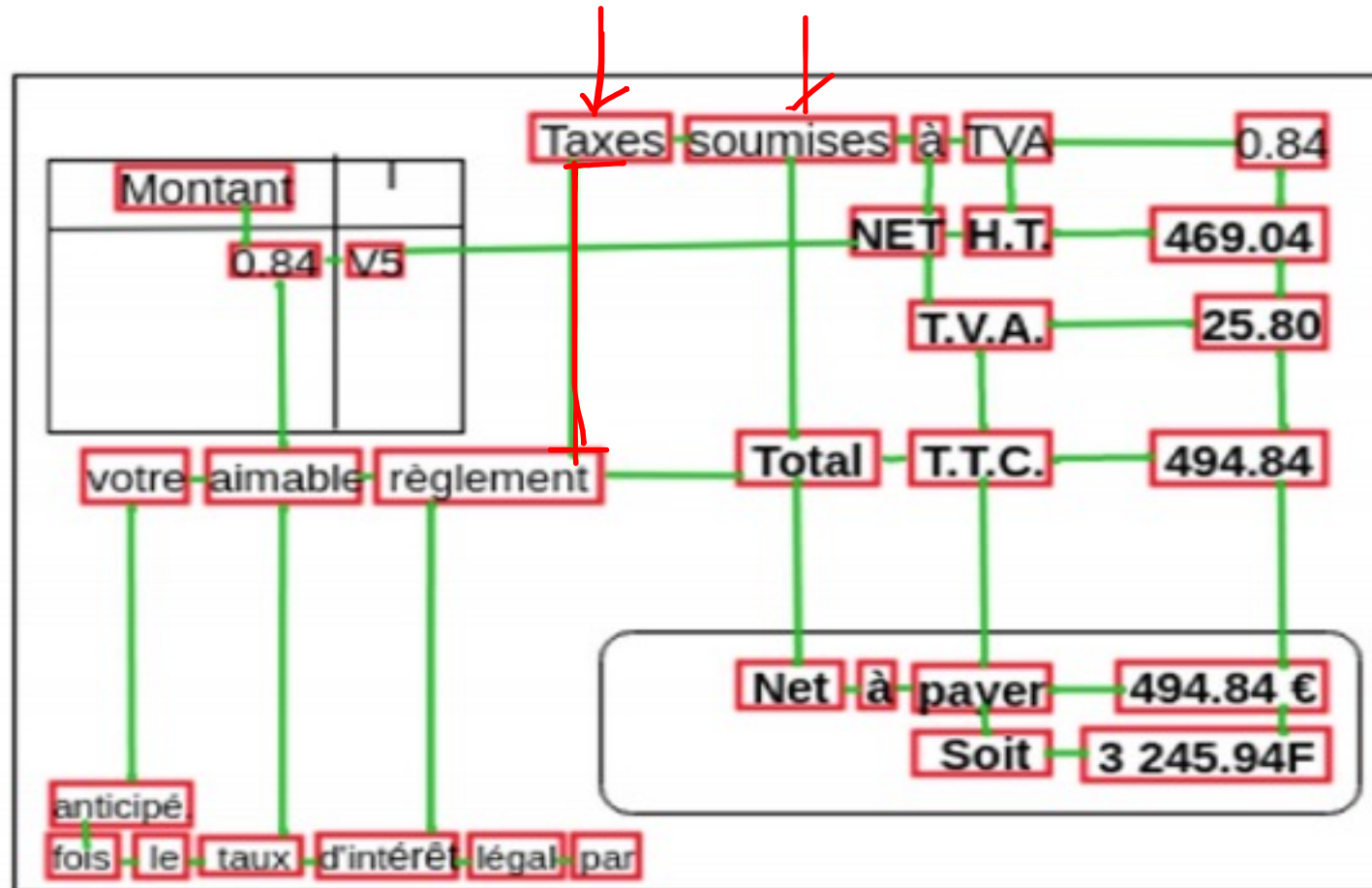
GCN
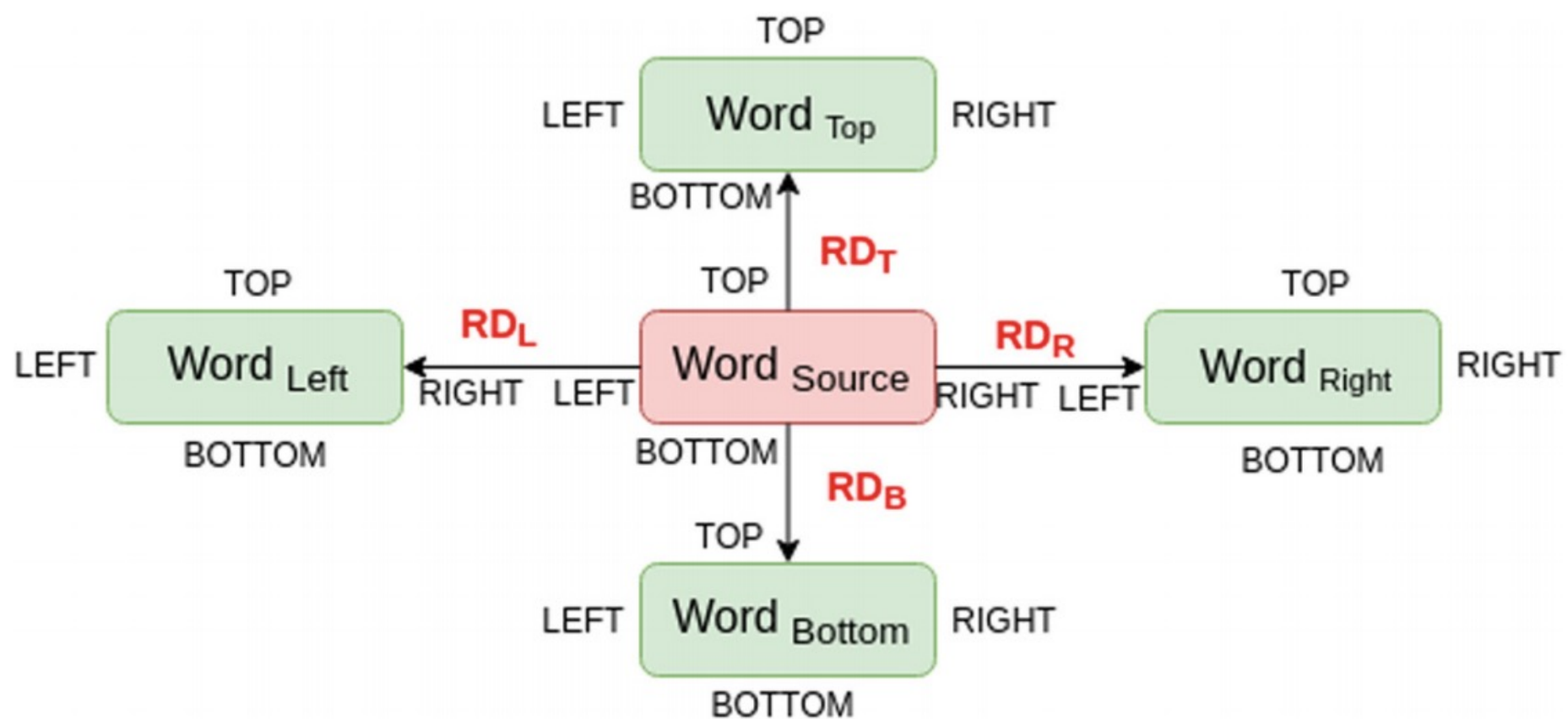
GCN

$$RD_L = (Right(Word_{Left}) - Left(Word_{Source}))/Width_{Page}$$

$$RD_T = (Bottom(Word_{Top}) - Top(Word_{Source}))/Height_{Page}$$

$$RD_R = (Left(Word_{Right}) - Right(Word_{Source}))/Width_{Page}$$

$$RD_B = (Top(Word_{Bottom}) - Bottom(Word_{Source}))/Height_{Page}$$

GCN

Text Segments of Document

Document Graph

Receipt 1
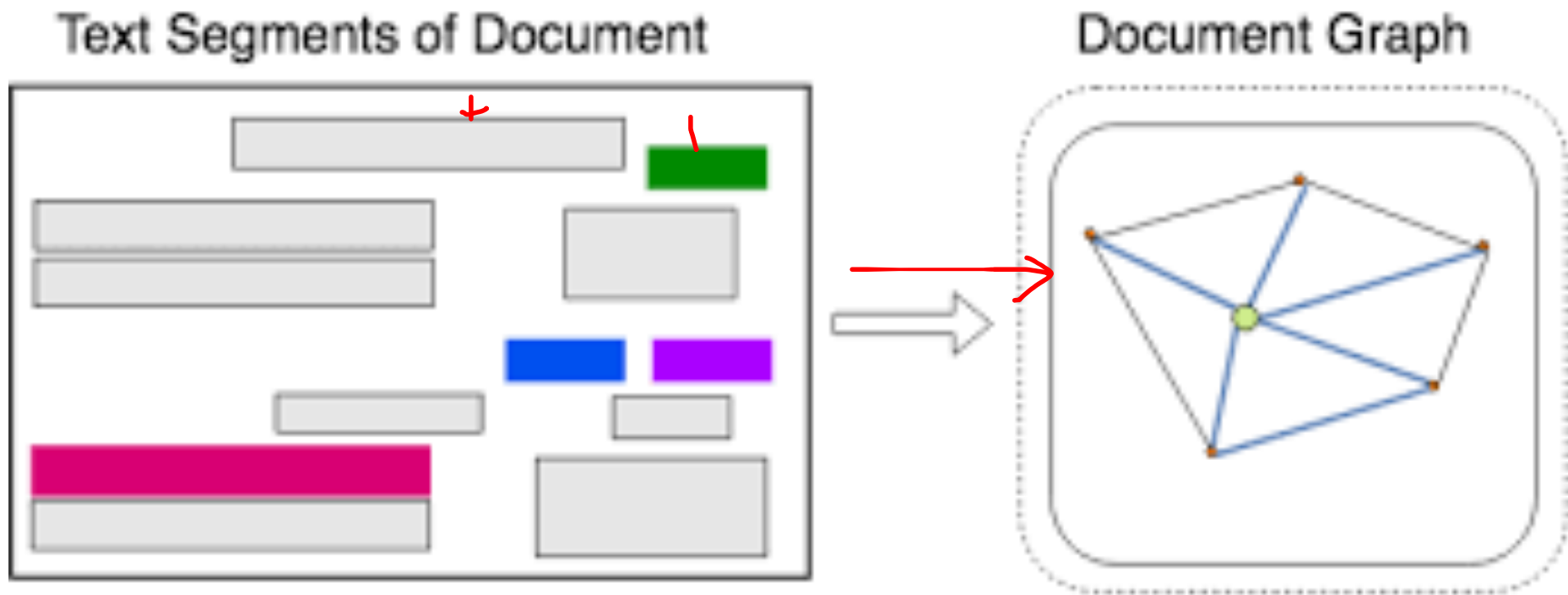
company ONE THREE SEAFOOD RESTAURANT SDN BHD
(1120908-M)
address NO.1, TAMAN SRI DENGKIL, JALAN AIR HITAM
address 43800 DENGKIL, SELANGOR.
(GST Reg. No : 000670224384)

Tax Invoice
Table 24
Bill#:V001-540401   Order#: 140508
Date : **date** 14-06-2018 20:41:44
Cashier: 113 CASHIER   Pax(s): 0

Qty	Description		Total TAX
1	Tilapia	D	28.00 SR
1	Vege item	D	16.00 SR
1	White Rice	D	7.00 SR
1	Beverage	D	5.00 SR
1	Herbal Tea Item	D	3.00 SR

Total (Excluding GST):   59.00
Total (Inclusive of GST):   59.00
total TOTAL :   59.00

Closed: 1   14-06-2018   21:22:27
Server: 113 CASHIER
CASH :   59.00

GST Summary	Amount(RM)	Tax(RM)
SR (@ 0%)	59.00	0.00

**** Thank You ****
Please Come Again

Receipt 2

company RESTORAN WAN SHENG
002043319-W
address No.2, Jalan Temenggung 19/9,
address Seksyen 9, Bandar Mahkota Cheras,
address 43200 Cheras Selangor
GST REG NO: 001335787520

Tax Invoice

INV No.: 1128507 Cashier: Thandar
Date   : **date** 06-05-2018 12:43:19

Description	Qty	U.price	Total TAX
Kopi (B)			
	1 x	2.20	2.20 SR
Take Away			
	1 x	0.20	0.20 SR

Total QTY: 2
Total (Excluding GST):   2.27
GST payable (6%):   0.13
Total (Inclusive of GST):   2.40
total TOTAL :   2.40

                                 2.40

GST Summary	Amount(RM)	Tax(RM)
SR (@ 6%)	2.27	0.13

Receipt 3

company KING'S CONFECTIONERY S/B 273500-U (KJ1)
address NO. 20-A1, JALAN SS6/3, KELANA JAYA,
address 47301 PETALING JAYA
address SELANGOR D.E.
TEL/FAX :- 03 - 78800732   TEL:0122144371
(GST ID : 000294060032)

Tax Invoice

Inv No :KJ130659712   **date** Date:09/02/17 11:12
Cashier:FW1254   Ctr :01

| Code	Desc	Qty	RM	D%	Amt
181010	French Lo	8	2.40	0	19.20 S

| | Qty | 8 | | | 19.20 |
| Rounding Adj | | | | | 0.00 |

Due   **total** 19.20
Pay   20.00

Change   0.80
===========================
(Total Included GST @ 6% : 1.09)
GOODS SOLD NOT RETURNABLE/EXCHANGEABLE!
TERIMA KASIH SILA DATANG LAGI!TQ
TASTY BAKING HEALTY NATION!!
==Visit us online at www.kings.com.my==

Handson

- **Planetoid** dataset/ https://graphsandnetworks.com/the-cora-dataset/

Planetoid

The citation network datasets `"Cora"`, `"CiteSeer"` and `"PubMed"` from the "Revisiting Semi-Supervised Learning with Graph Embeddings" paper.

The Cora dataset consists of 2708 scientific publications classified into one of seven classes. The citation network consists of 5429 links. Each publication in the dataset is described by a 0/1-valued word vector indicating the absence/presence of the corresponding word from the dictionary. The dictionary consists of 1433 unique words.

Handson

	TARGET	SOURCE	LABEL
3581	72908	93923	cites
5303	656231	103531	cites
2005	14531	592830	cites
987	4330	37884	cites
1695	10183	1120713	cites

	W_0	W_1	W_2	W_3	W_4	W_5	W_6	W_7	W_8	W_9	...	W_1424	W_1425	W_1426
31336	0	0	0	0	0	0	0	0	0	0	...	0	0	1
1061127	0	0	0	0	0	0	0	0	0	0	...	0	1	0
1106406	0	0	0	0	0	0	0	0	0	0	...	0	0	0
13195	0	0	0	0	0	0	0	0	0	0	...	0	0	0
37879	0	0	0	0	0	0	0	0	0	0	...	0	0	0

```
set(node_data["subject"])
{'Case_Based',
 'Genetic_Algorithms',
 'Neural_Networks',
 'Probabilistic_Methods',
 'Reinforcement_Learning',
 'Rule_Learning',
 'Theory'}
```

Handson

Node Classification with GNN

We will create a GCN model structure that contains two **GCNConv** layers **relu** activation and a **dropout** rate of 0.5. The model consists of 16 hidden channels.