

Graph Neural Network for Named Entity Recognition and Nested Named Entity Classification

Presented by:

Anas OUARDINI

01/09/2023

Supervised:

Pr. **LUDOVIC MONCLA**

Outline

- Introduction
- Objectif of Study
- Data Preprocessing
- Experiments and results
- Conclusion

Introduction

Named Entity Recognition and Classification (NERC) using Graph Neural Networks (GNNs).

ORGANISATION

LOCATION

DATE

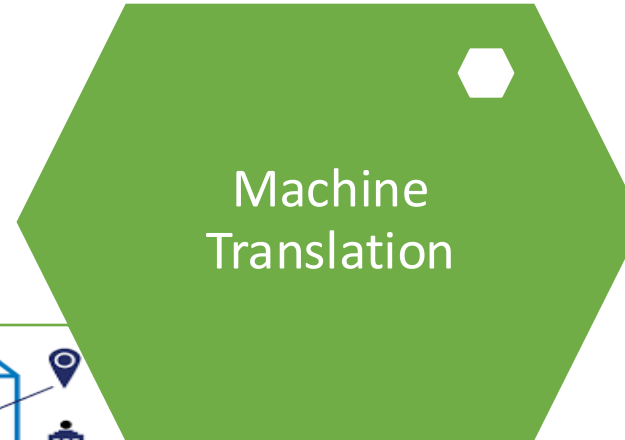
PERSON

WEAPON

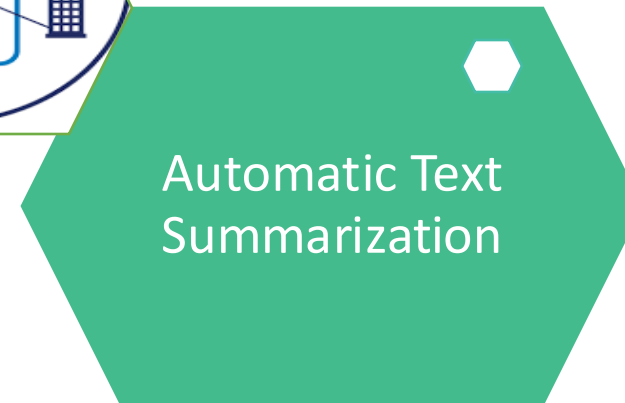
The **ISIS** ORG has claimed responsibility for a suicide bomb blast in the **Tunisian** LOC capital **earlier this week** DATE, the **militant group** ORG 's **Amaq news agency** ORG said on **Thursday** DATE. A **militant** PER wearing an **explosives belt** WEAPON blew himself up in **Tunis** LOC

Example

ML for NER



ML for NNE



Objective of the Study

To enhance NERC using GNNs and address challenges in handling complex named entities.

Highlight the significance of achieving accurate named entity recognition in language understanding tasks.

Challenges in Named Entity Recognition



Highlight the challenges in traditional NERC methods, especially when dealing with nested structures and contextual dependencies.

Emphasize the limitations of rule-based and dictionary-based approaches.

Dictionary-based

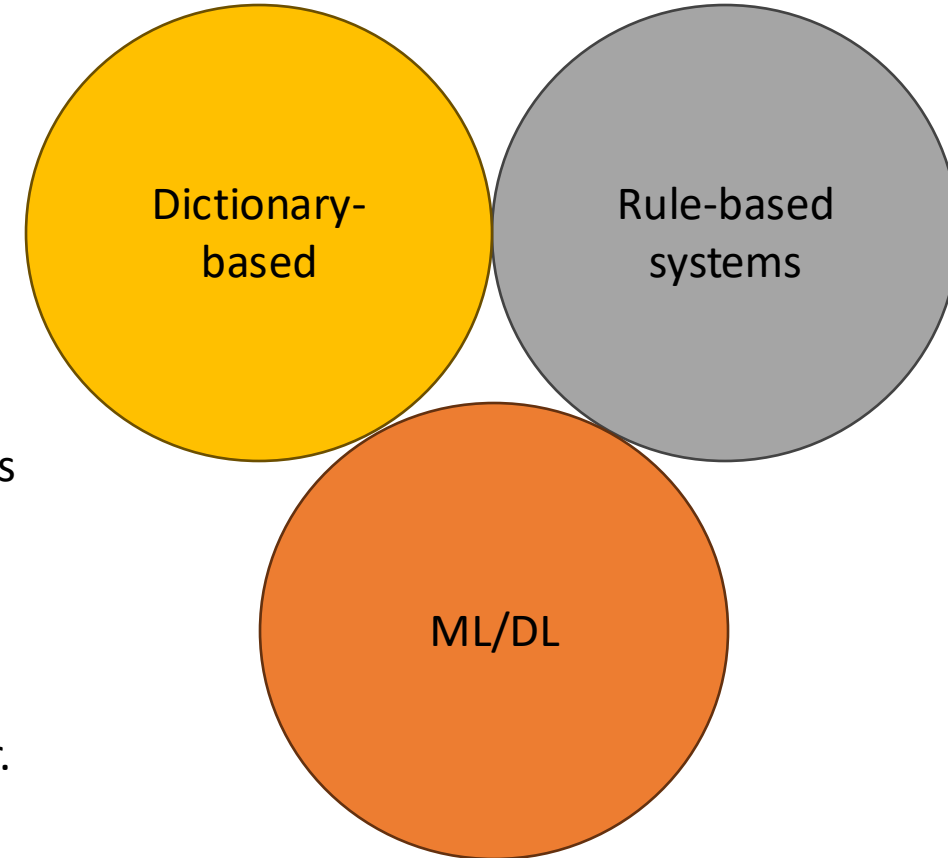
This is the simplest NER method. In this approach, a dictionary containing vocabulary is used.

Rule-based systems

Here, the model uses a pre-defined set of rules for information extraction. This type is used context-based rules, which depend upon the context of the word used in the given text document.

NER and ML/DL

In this model, a set of training data is used to train the classifier and a set of unlabeled data (test data) is used to evaluate the performance of the classifier.



Experiment (DATA)

volume01-1465.txt - Bloc-notes

Fichier Edition Format Affichage Aide
AISAY-LE-DUC, (Géog.) ville de France en
Bourgogne, Bailliage de Chatillon.

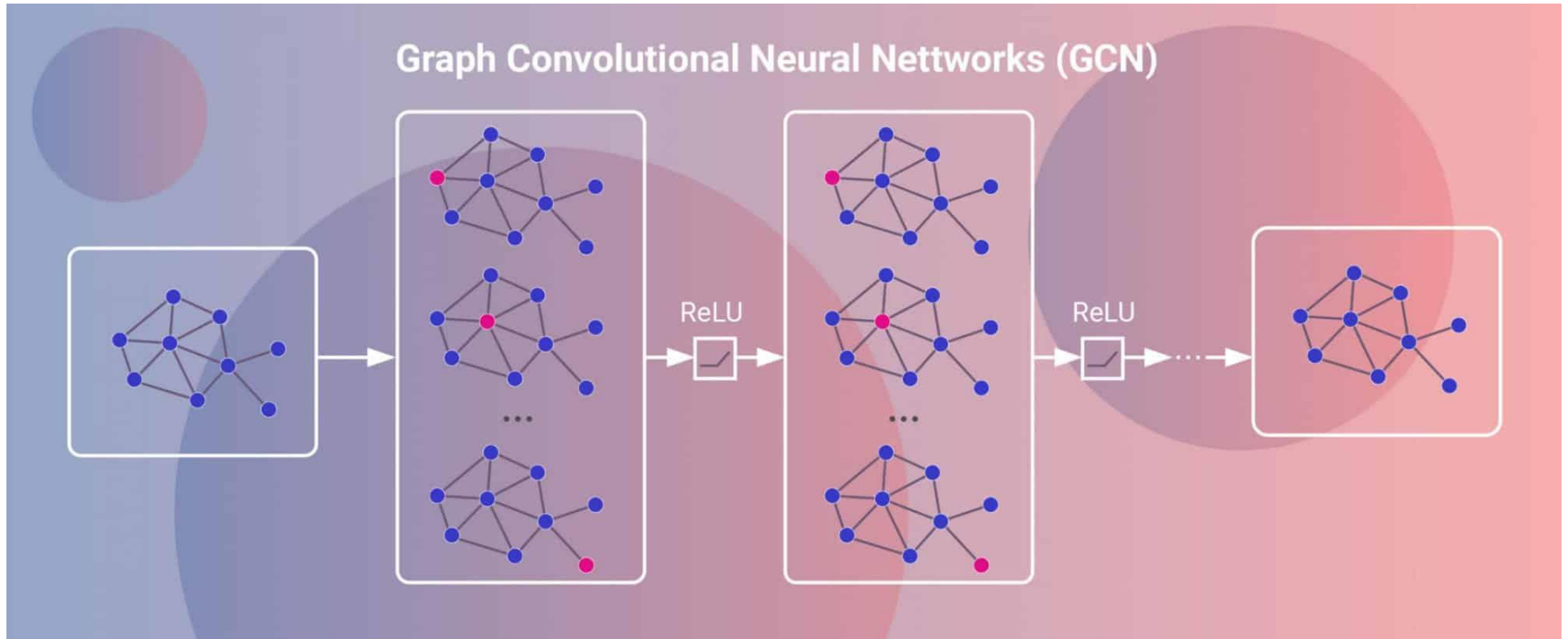
By Perdido OR
Manually method

Gold Standard Corpora (GSC) is a small annotations data are performed manually.

Silver Standard Corpora (SSC) is an annotations are automatically.

```
▼<rs type="ENE_Spatial">
  ▼<rs type="ENE_Spatial">
    ▼<rs type="ENE_Spatial">
      ▼<rs type="NC_Spatial">
        <w id="1-7" b_span="22" e_span="27" pos="N" lemma="ville">ville</w>
      </rs>
      ▼<rs type="Relation">
        <w id="1-8" b_span="28" e_span="30" pos="PREP" lemma="de">de</w>
      </rs>
      ▼<rs type="NP_Spatial">
        <w id="1-9" b_span="31" e_span="37" pos="NPr" lemma="france">France</w>
      </rs>
      </rs>
      ▼<rs type="Relation">
        <w id="1-10" b_span="38" e_span="40" pos="PREP" lemma="en">en</w>
      </rs>
      ▼<rs type="NP_Spatial">
        <w id="1-11" b_span="41" e_span="50" pos="NPr" lemma="bourgogne">Bourgogne</w>
      </rs>
      </rs>
      <w id="1-12" b_span="50" e_span="51" pos="PUN" lemma=",">,</w>
    </rs>
    ▼<rs type="ENE_Spatial">
      ▼<rs type="NC_Spatial">
        <w id="1-13" b_span="52" e_span="61" pos="N" lemma="bailliage">Bailliage</w>
      </rs>
      ▼<rs type="Relation">
        <w id="1-14" b_span="62" e_span="64" pos="PREP" lemma="de">de</w>
      </rs>
      ▼<rs type="NP_Spatial">
        <w id="1-15" b_span="65" e_span="74" pos="NPr" lemma="chatillon">Chatillon</w>
      </rs>
      </rs>
      <w id="1-16" b_span="74" e_span="75" pos="SEN" lemma=".">.</w>
    </rs>
  </rs>
</rs>
```


GNNs and GCNs : Overview



Research Methodology

(1) Annotations (Automatique et manuel)



ville	Place
de	O
France	Place
en	O
Bailliage	Place
,	O
Baillage	Place
de	O
Chatillon	Place
.	O

(2) Extraction des caractéristiques

POS Tag
Étiquette du mot
FastText (features des nodes)
Poids sur des arêtes
....

(3) Entraînement du classifieur



Modèle NER



.....

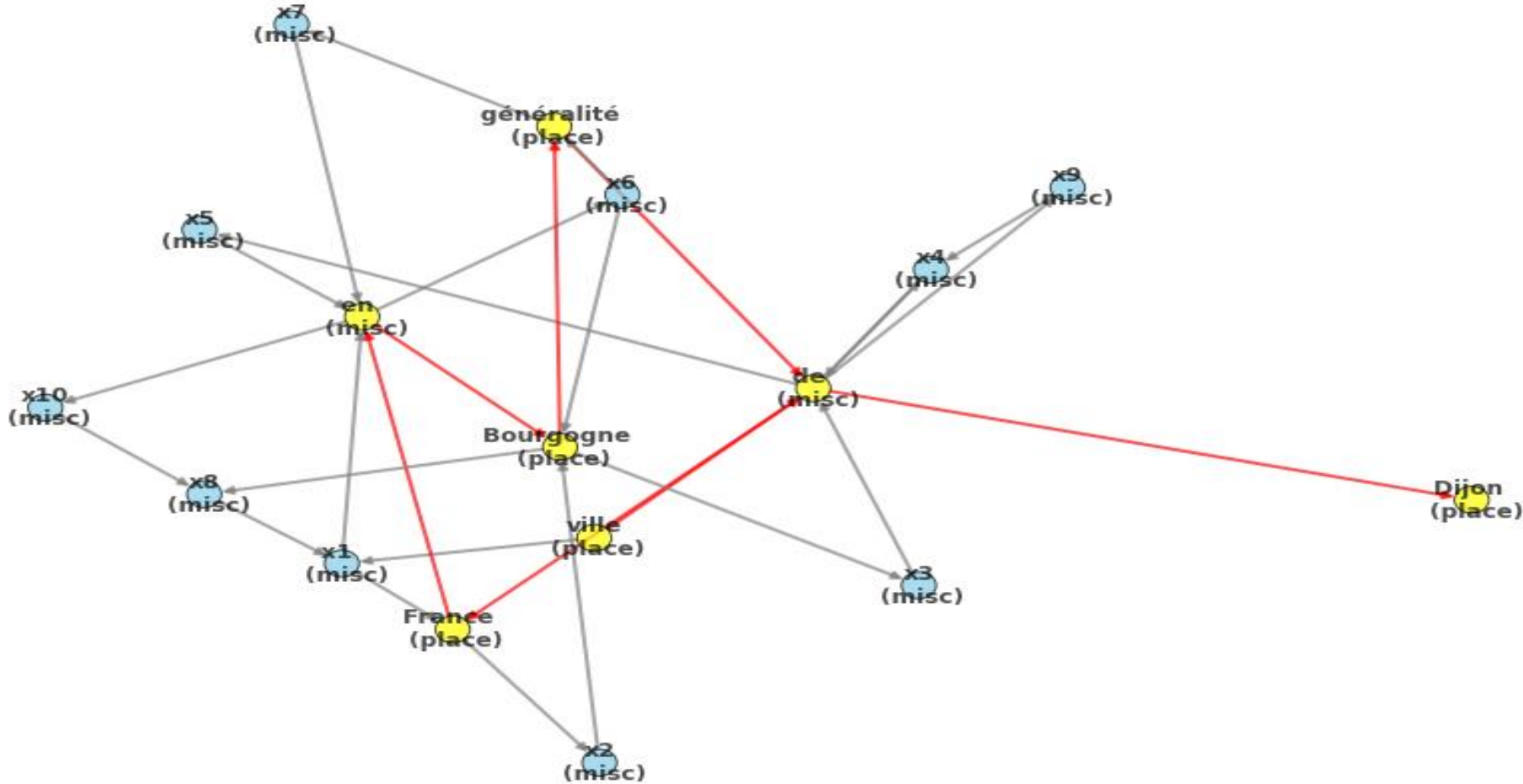
Classification

- **Node classification:** The task here is to predict the labeling of nodes by looking at the labels of their neighbors and the features of this nodes.

- **Graph classification:** We are focused on predicting labels for subgraphs, determining whether they represent a place, a person, or another entity.

. Node Classification for NERC

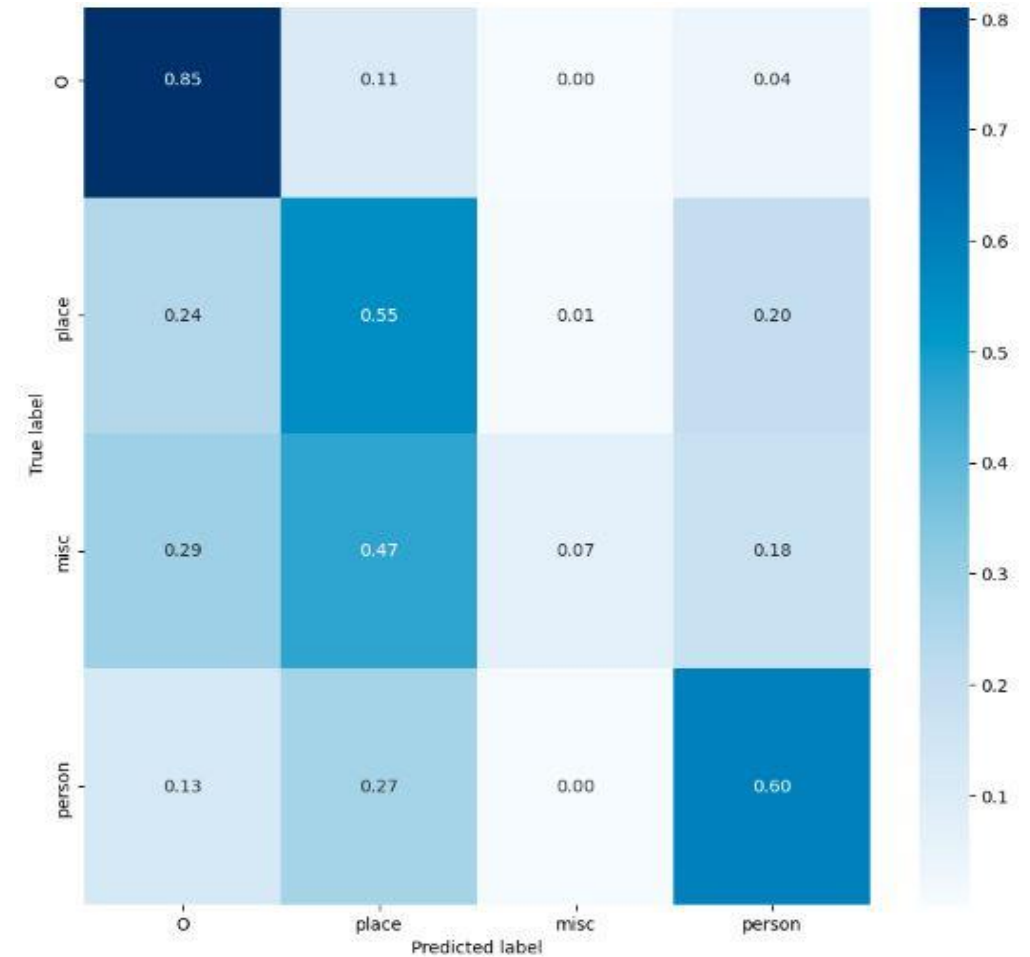
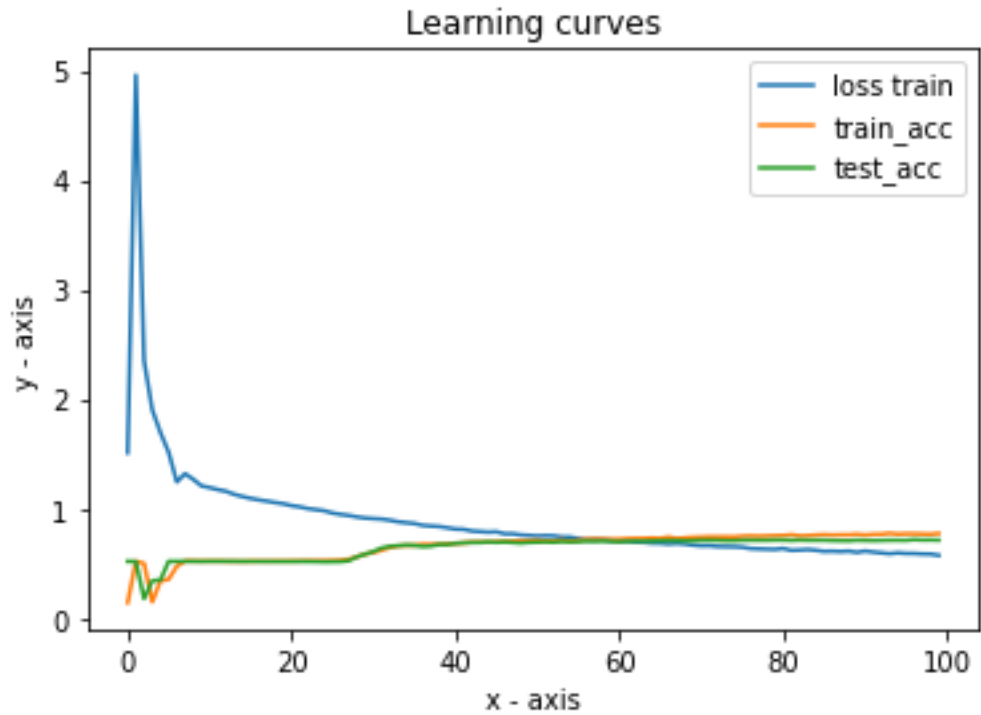
Example of Corpus: Directed Graph Visualization with Colored Nodes and Edges



- **Node classification:** used (DGL and PyTorch-Geometric) on the two types of data (manual and automatic).

RESULTS

Manual Data for training and testing



Node Classification Results

Manual Data for training and testing

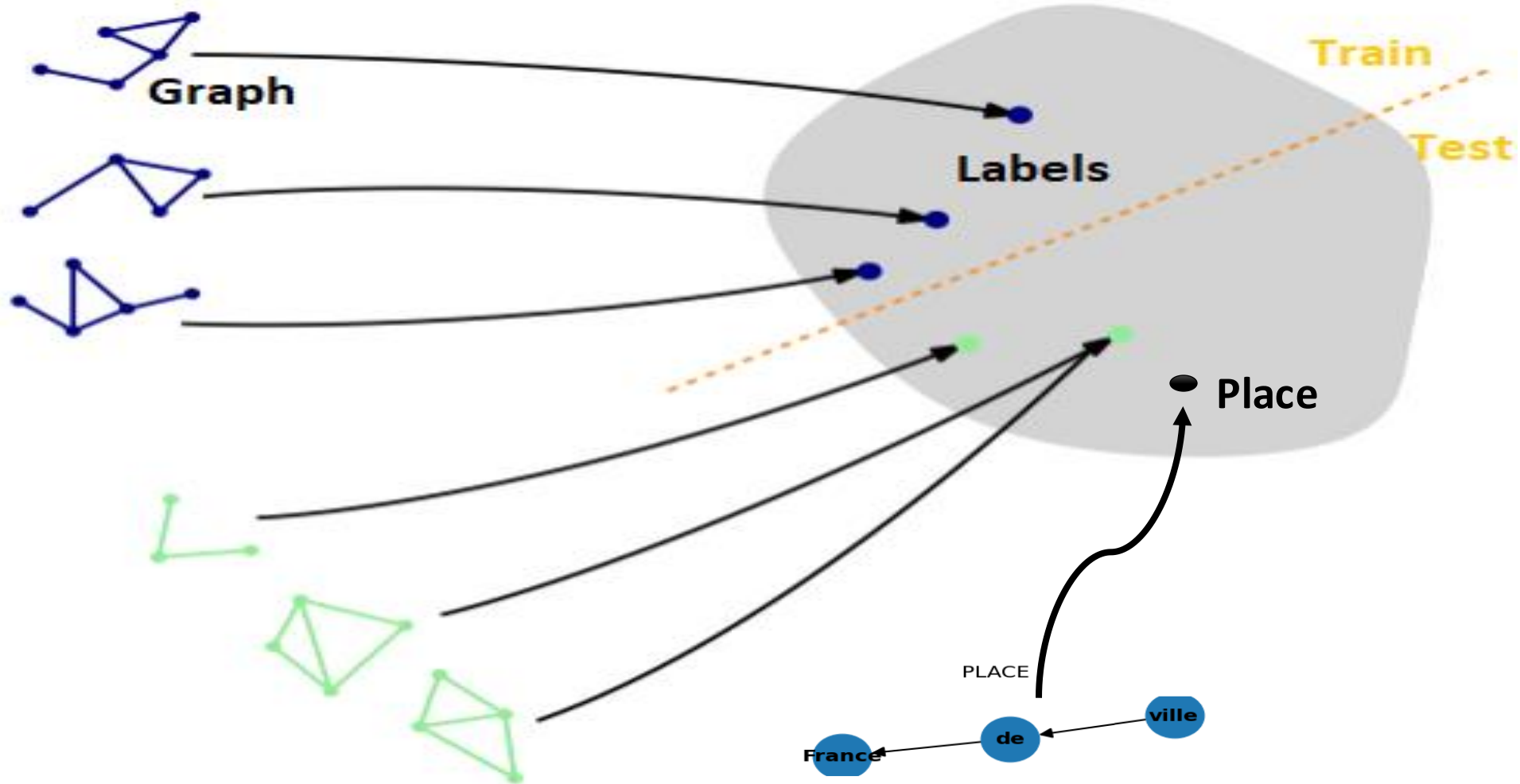
Type of Data	Modes			
	DGL	Pytorch	Bert	Perdido
Data	70%	68%		
Graph	71%	69%	71%	70%

Classification

- **Node classification:** The task here is to predict the labeling of nodes by looking at the labels of their neighbors and the features of this nodes.

- **Graph classification:** We are focused on predicting labels for subgraphs, determining whether they represent a place, a person, or another entity.

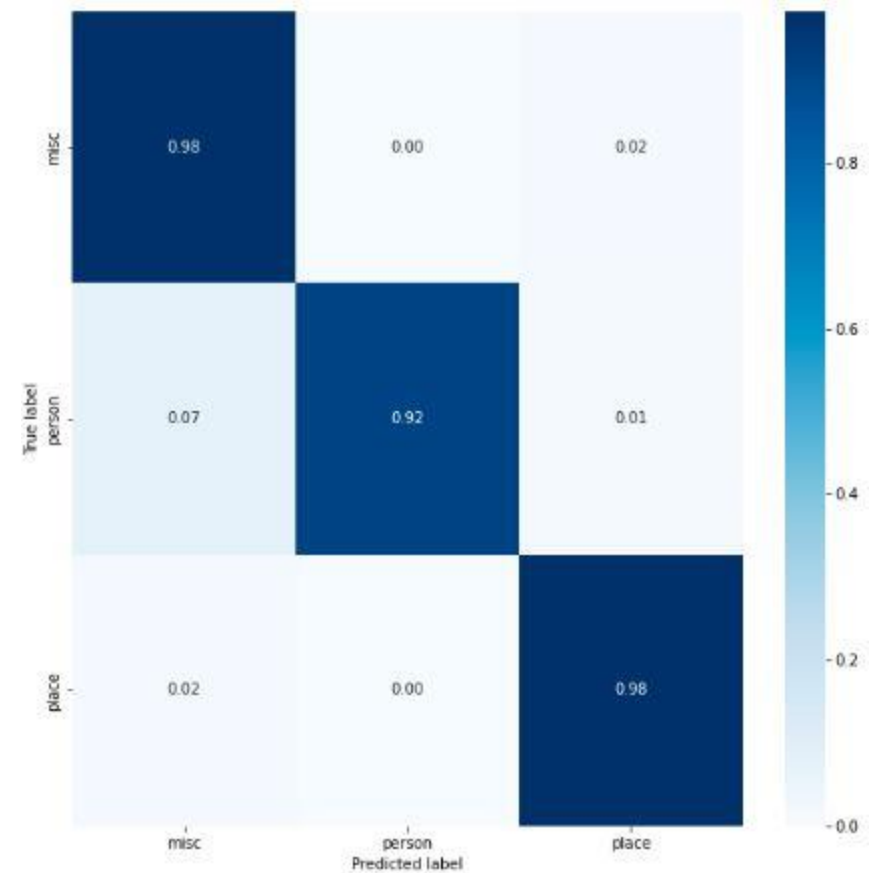
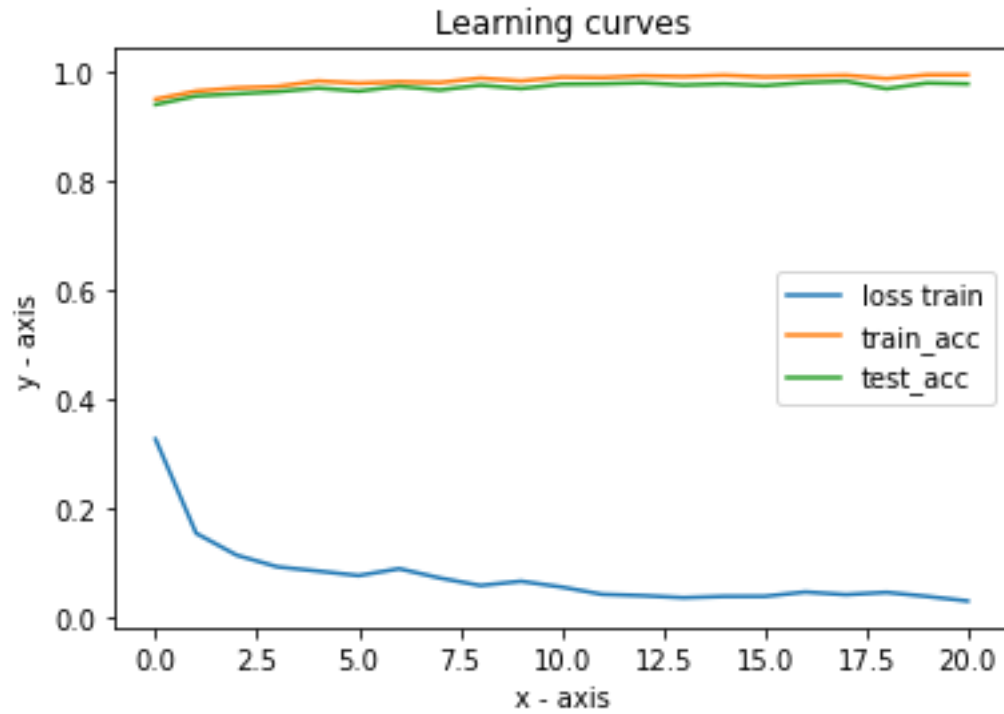
Graph Classification for Nested Named Entity Classification (NNEC)



- **Graph classification:** used (DGL and PyTorch-Geometric) on the two types of data (manual and automatic).

Graph Classification Results

Manual Data for training and testing



Graph Classification Results

Manual Data for training and testing

Type of Data	Models				
	DGL	Pytorch	Bert	Transformer	Perdido
Data	70%	98%	98%	71%	89%
Graph	85%	88%			

Conclusions & Future Directions

To summarize, our study underscores the transformative capability of employing GNNs in conjunction with PyTorch for advancing Named Entity Recognition and Classification (NERC).

Thank you for your attention
