IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Authors
Gabriel GROS
Laurine MARTY

Supervisor : Vincent GRIPON

Affiliates
Resarch program
IMT Atlantique
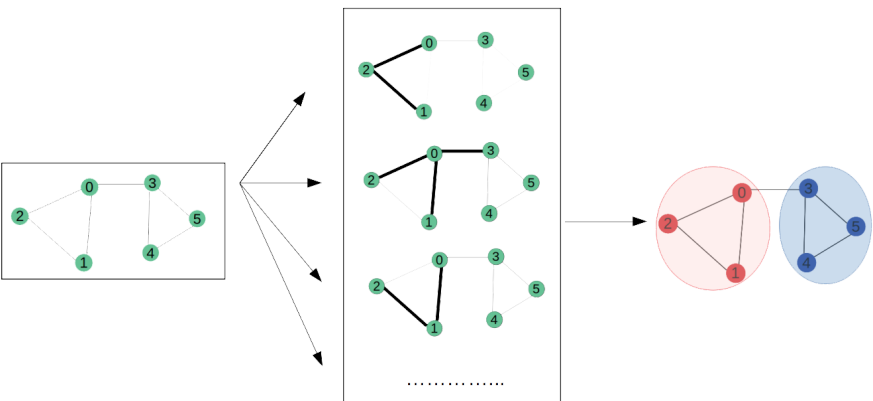Lab STICC

Lab·STICC

Graph Convolutional Networks

Introduction

- Traditionally: Deep Learning datasets represented in the Euclidean space.
- But recently: increasing number of non-Euclidean data represented as graphs.
- Need for a new type of convolutional neural network: GCN (Graph Convolutional Network) which is a neural network operating on graphs.

Various applications: Natural Language Processing, Traffic control, Chemistry, etc.

Objective

IMT Atlantique laboratories are getting old and owe their organization to historical rather than technical reasons. However, thinking about a new organization is time consuming. Why not think about a new distribution of the school researchers thanks to Deep Learning on graphs?

GCN Approach

GCN has to transform the graph :

- Enforcing self-connections by adding the identity matrix to the adjacency matrix A.
- Symmetric normalization by degree matrices

Then, GCN updates each node's vector aggregating its neighbors'information.
For all nodes of G:

- Get features of neighbors
- Update node feature

$$h_{v_i}^{(l+1)} = \sigma \left(\sum_j \frac{1}{c_{ij}} h_{v_j}^{(l)} W^{(l)} \right)$$

HOW IS THE GRAPH OBTAINED ?

To create the graph, we had to look at HAL database to select IMT Atlantique researchers. The objective is to be able to enter the first crews so that the GCN can match or not reasearchers in laboratories.

We then had to create a dataset including:

- The author ID.
- His co-authors.
- His papers' titles and abstracts.
- Eventually, the laboratory is has been attributed to.
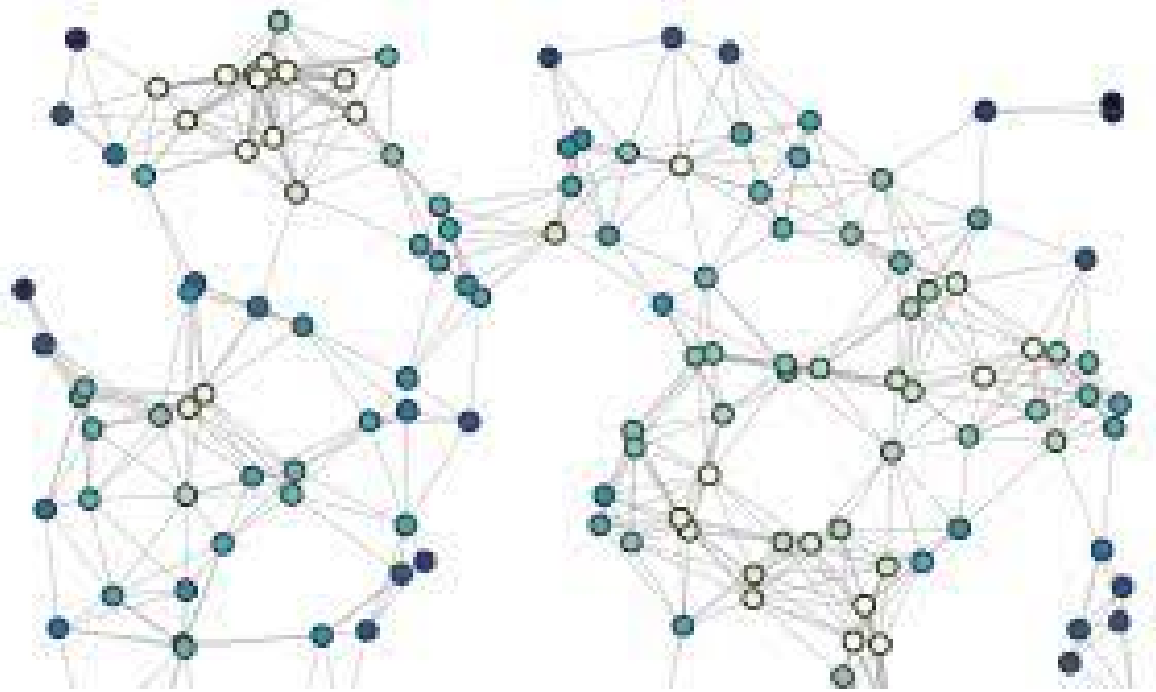
Our Dataset

Has to be inspired by the CORA dataset: graph of scientific articles that cite each other

The aim is to build a graph of IMT Atlantique co-authors with :

- Each node : one hot vector representing the author from his papers' vocabulary
- Each edge : link between two authors who wrote at least once together

The graph: symmetrical sparse matrix

CSV transform

Name	Id	transf_Name	Classes	Co_auth	Attribute
Cyril Lahuec	17541	cyrillahuec	0		3.0
Alexandre Dolgui	30378	alexandredolgui	5		657265.0
Sophie Bretesché Ay	53912	sophiebretesche	0		3.0
Thibault de Swarte	55066	thibaultdeswarte	0		3.0
Romain Billot	59488	romainbillot	5		5.0
Yannis Haralambous	62300	yannisharalambous	0	896137,896137	5.0
Philippe Lenca	65458	philippelenca	0		738665 3.0
Sorin Moga	65726	sorinmoga	0		303785.0
John Puentes	66545	johnpuentes	0		465196 9.0

Results

After a training phase, the GCN returns a matrix of probability vectors associated with each class. A certain index corresponds to the associated probability vector.

Conclusion

- Laboratories attribution is the typical case of a problem which can't be solve only with the classical mathematical object.
- Graphs have allowed us to take into account both the work of the researchers and the links they have with the community.
- The strength of this method is that it takes into account the links between objects during training even though they have not been labelled .
- The lack of an optimal solution for this problem prevents us from objectively quantifying its performance

TARGET NODE

INPUT GRAPH