# Edge Existence Prediction Based on Node Embedding

**Team Members:** Jintian Ge, Binyu Xia, Roushuang Ye
**Instructor:** Prof. Yifei Zhu

## Overview

**Goal**: given two groups and some edges between these two groups, predict the rest of the potential edges between these groups.
**Solution**: Using node embedding method to calculate attributes for each node, then use deep learning algorithm to predict the connection.
**Challenges**: How to determine attributes. How to construct a valid learning algorithm.

## Motivation

Nowadays, relationships between people and people become more and more important. Sometimes, if there are two groups, we may be interested in whether these two groups can be connected in some ways. For example, assume that there are two classes in JI, class No.1 and class No.2. We may want to find out some potential friendships between students in class 1 and students in class 2 to help introduce new friends.

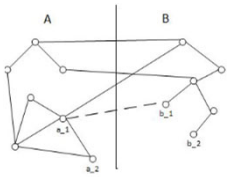


**Fig.1** Friendship Prediction

## Problem Statement

Given two isolated groups (most connection between the nodes from each group is unknown), we want to predict whether an edge exist given two nodes from one of each groups. Namely, we want an algorithms as following:

**Algorithms:** *Merging Algorithms*

***Input***: $\boldsymbol{Net_1, Net_2}$:Network to predict, $\boldsymbol{C_i(v_1, v_2)}$:The known connectivity between some nodes from $Net_1, Net_2$.

***Output***: $\boldsymbol{C_i(v_1, v_2)}$: The predicted connectivity between the rest nodes from $Net_1, Net_2$.

## Algorithm Design

We will use node embedding algorithm to estimate vertices features in two networks. Then, we use edges.

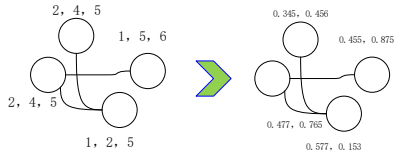


**Fig.2** Node Embedding

We between these two networks to train our prediction machine learning algorithm, which is a neural network. Finally, we will use the trained machine learning algorithm to predict whether there is a potential edge between two vertex in these two networks. If so, we will link them together. By repeating this process, finally we can merge two networks together.

1. *Embedding* $Net_1$
2. *Embedding* $Net_2$
3. *Training* $Net_1$, $Net_2$ *with known Edges*
4. *Predict the unknown edge existence with the training results*

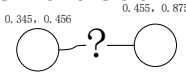


**Fig.3** Edge Prediction

## Model Analysis

Our model is divided into two parts: Node embedding part and edge prediction part.

For edge prediction part, we use a four layers neural network. The first two layers are separated, while the last two layers are fully connected. This means that we will process input attributes of the two nodes separately, then make a prediction together.

For node embedding part, we use the same idea as the collaborative filtering to convert a sparse features graph into a dense one.

## Results

*Dataset description:*

We use the data set from Facebook. This web graph is a page-page graph of varied Facebook sites. Nodes represent social Facebook pages while the links are mutual likes between sites.

| | |
|---|---|
| **Directed:** | No. |
| **Node Features:** | Yes. |
| **Edge Features:** | No. |
| **Node Labels:** | Yes. Binary-Labeled |
| **Temporal:** | No. |
| **Nodes:** | 22,470 |
| **Edges:** | 171,002 |
| **Density:** | 0.001 |

**Compared Approaches:**

- "Line" Embedding: the method is suitable for arbitrary types of networks, including undirected, directed, weighted and so on.
- Collaborative Filtering Embedding: Our provides a methods to deal with the features of nodes based on our knowledge of network.

*Evaluation results:*

**Training Results Using Our Methods**

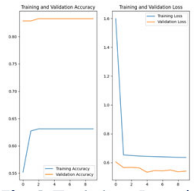

**Fig.3** Training Results 1

**Training Results Using Original Features**

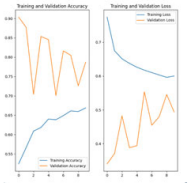

**Fig.2** Training Results 2

1. The results show that (1) our edge prediction model is valid; (2) our node embedding method is much stable than using original features directly.

## References

1. Mikolov, T., Chen, K., Corrado, G., and Dean, J. Ecient estimation of word representations in vector space. In International Conference on Learning Representations, Workshop Track Proceedings, 2013.
2. Liao, L., He, X., Zhang, H. and Chua, T. Attributed social network embedding. IEEE Transactions on Knowledge and Data Engineering, 20(12): 2257-2270, 2018.