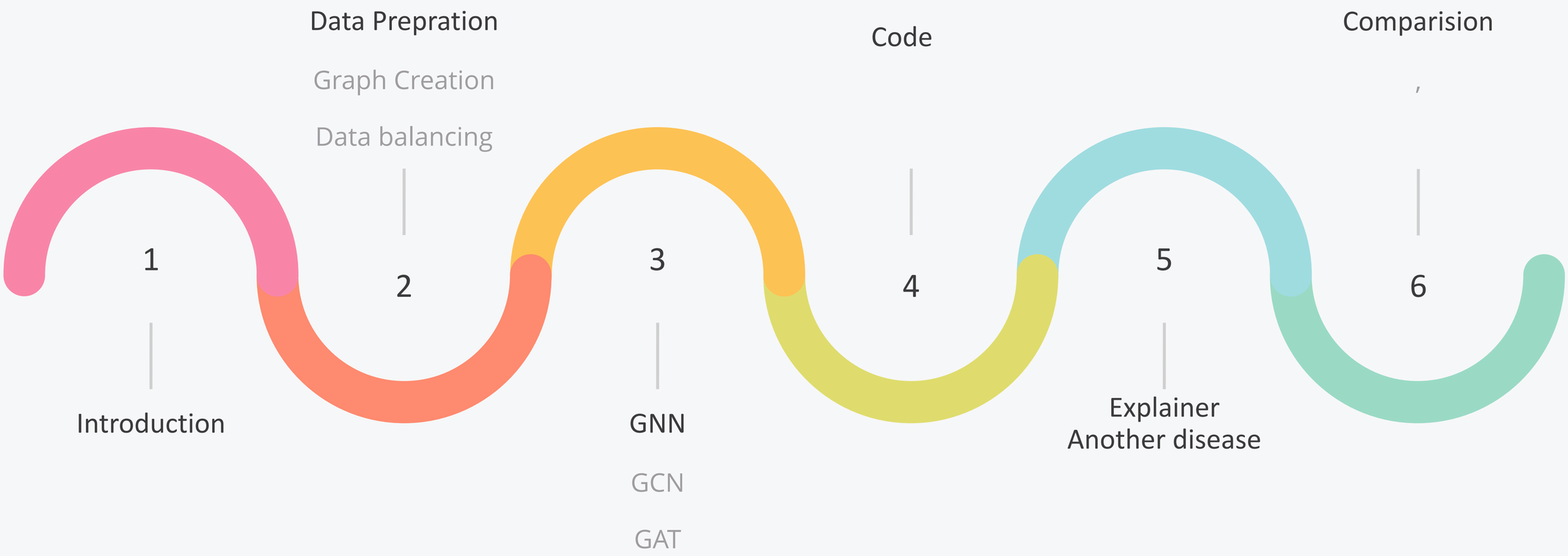
Gene Classification using GNNs

Presented by:

Alireza Samadifardheris

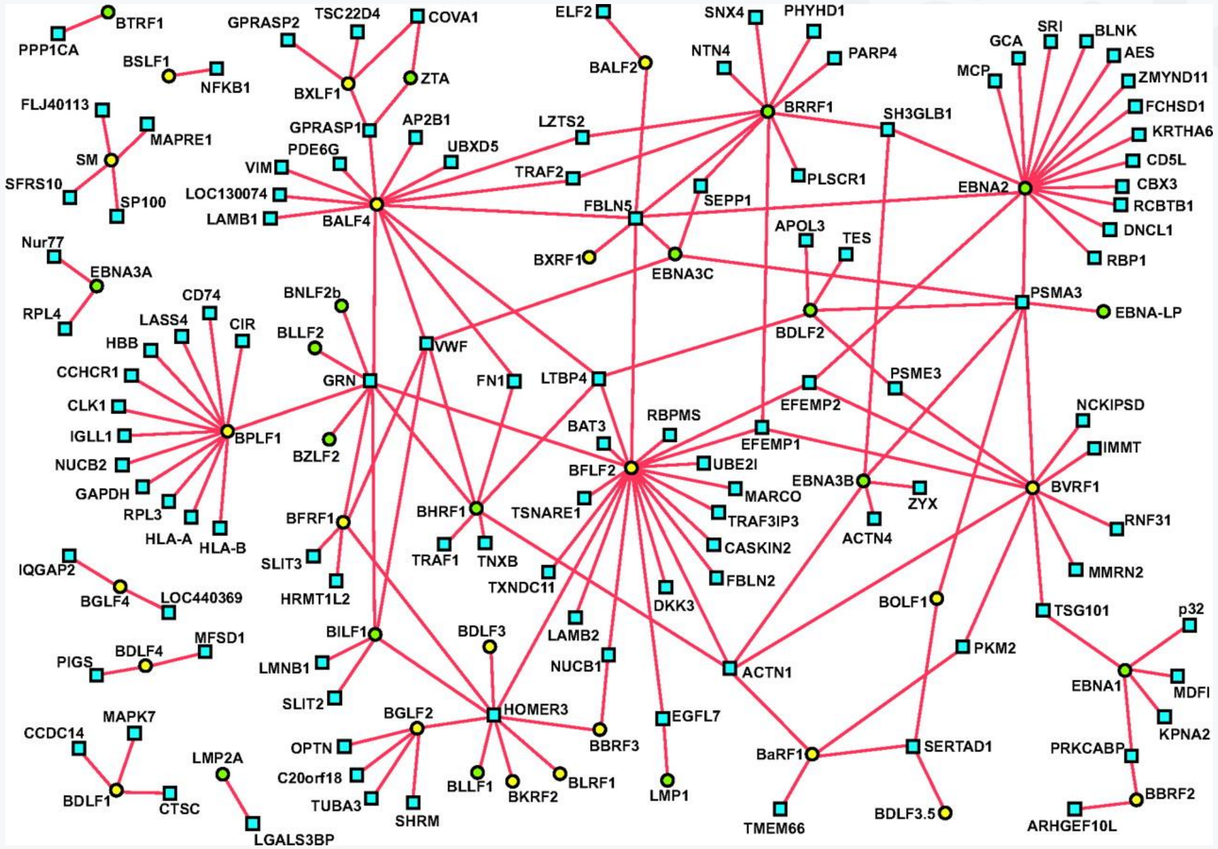
Road map


Data Prepration
Graph Creation
Data balancing
Code
Comparision
,
1
2
3
4
5
6
Introduction
GNN
GCN
GAT
Explainer
Another disease

1. Introduction

Provided data : bio grid file + gene-disease association
Task: node classification
Method:GNN models
Chosen models :
GCN
GAT
GraphSage

What it may look like :



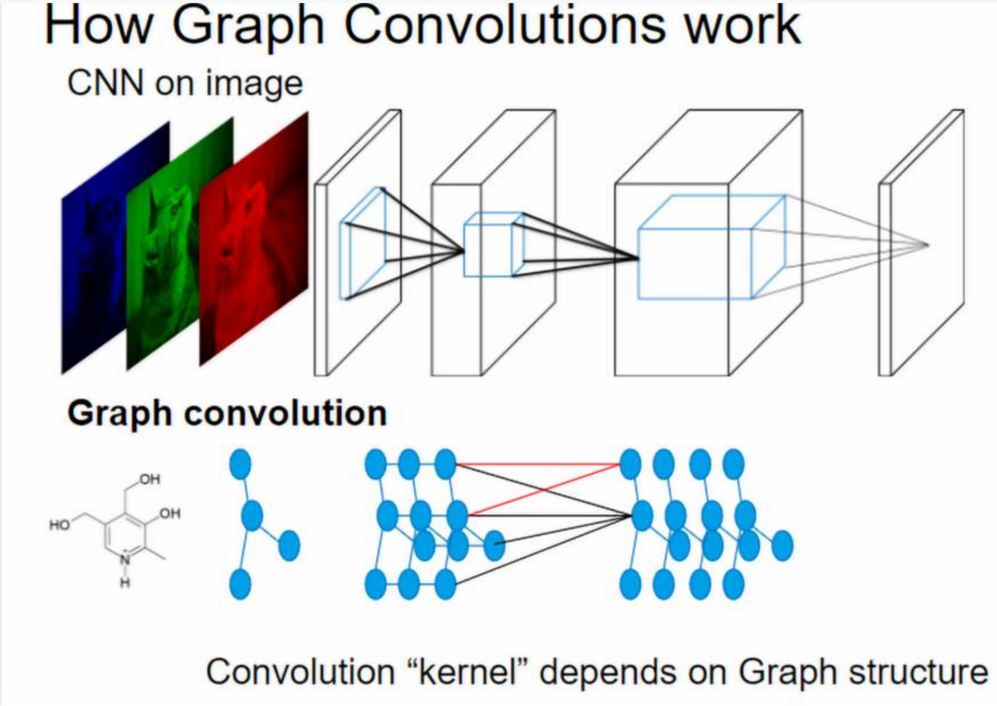
2. Data preparation

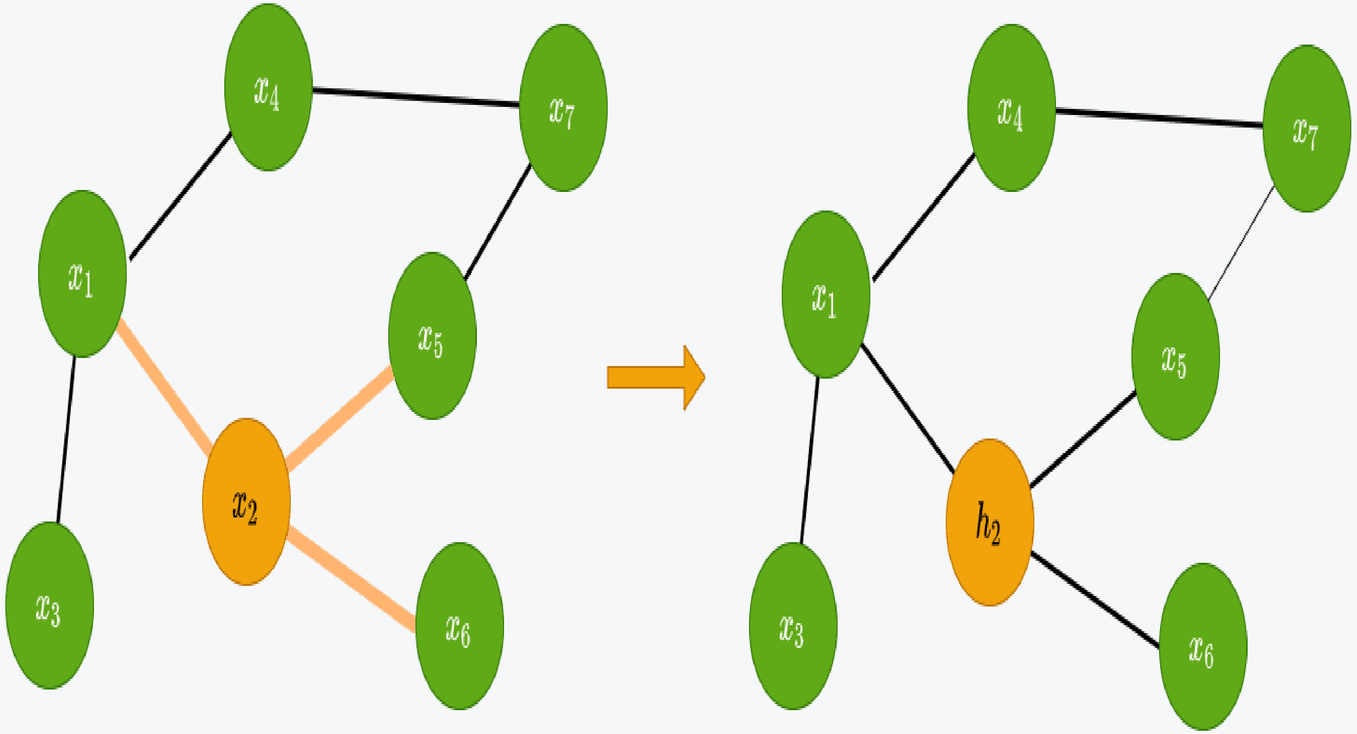
1. Chosing data(columns) we need from all_gene_disease_association (it’s a large file)
2. Chose disease: two different diseases to compare results
3. Create graph and labeling nodes
4. Consider if the graph is unbalanced: create a balanced version of graph

3.GNN

1. GCN (Graph Convolutional Neural network)
2. GAT (Graph Attention Network)
3. GraphSAGE

3.1. GCN





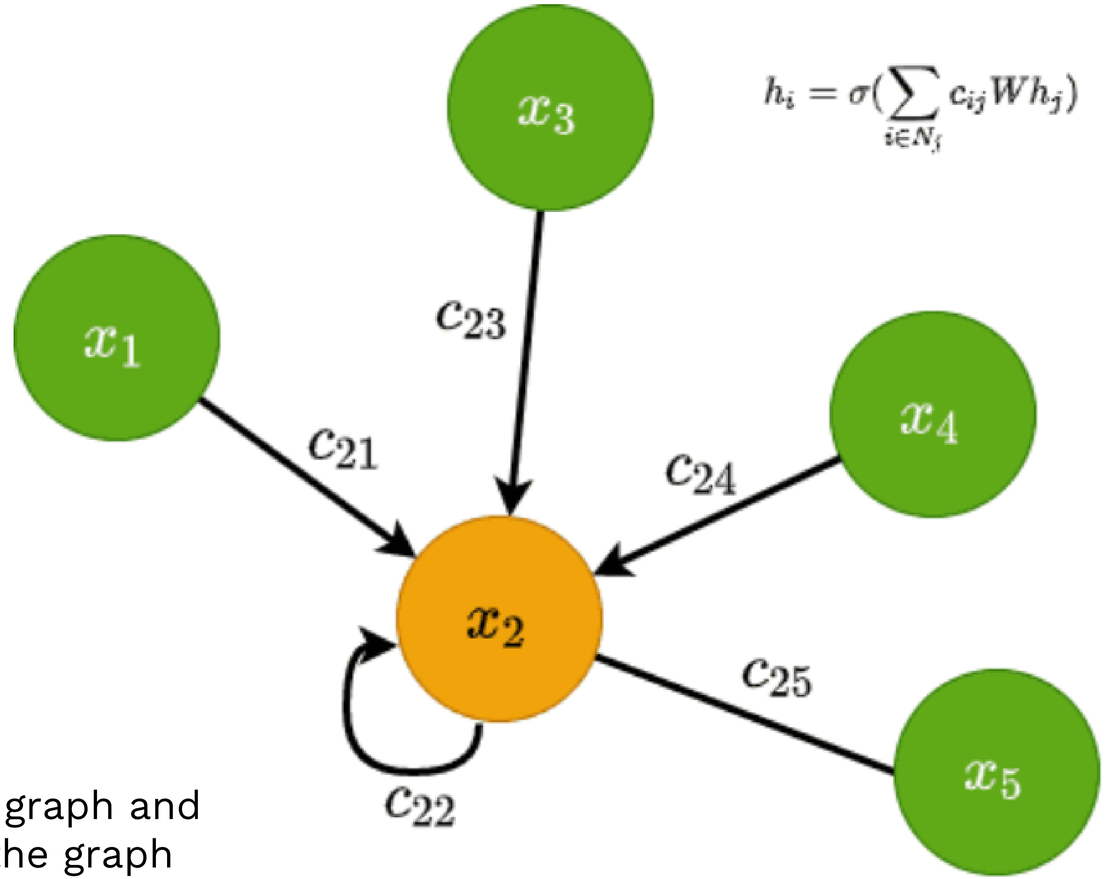
$$h_2 = g(x_1, x_5, x_6)$$

we try to generalize the idea of convolution into graphs
generalization of images where:
every node corresponds to a pixel connected to 8 (or 4) adjacent neighbors

From a node-wise perspective, the update rule can be written as :

$$h_i^{(l)} = \sigma(\sum_{i \in N_j} c_{ij} W h_j)$$

Where $c_{ij} = \frac{1}{\sqrt{|N_i||N_j|}}$, and N_i and N_j are the sizes of the nodes' neighbourhoods.



ci s derived from the degree matrix of the graph and is heavily dependent on the structure of the graph

3.1. GAT

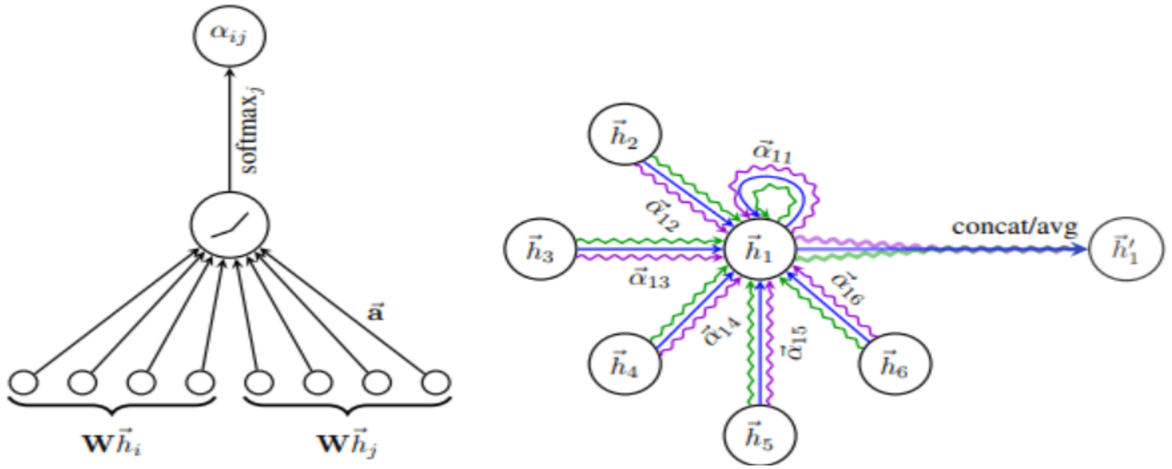
compute that coefficient implicitly rather than explicitly as GCNs do

the softmax function is applied in the attention weights a_{ij} to that result in a probability distribution. Mathematically we have:

$$a_{ij} = attention(h_i, h_j)$$

$$a_{ij} = \frac{exp(a_{ij})}{\sum_{k \in N_i} exp(a_{ik})}$$

Visually this can be seen on the left side of the following image

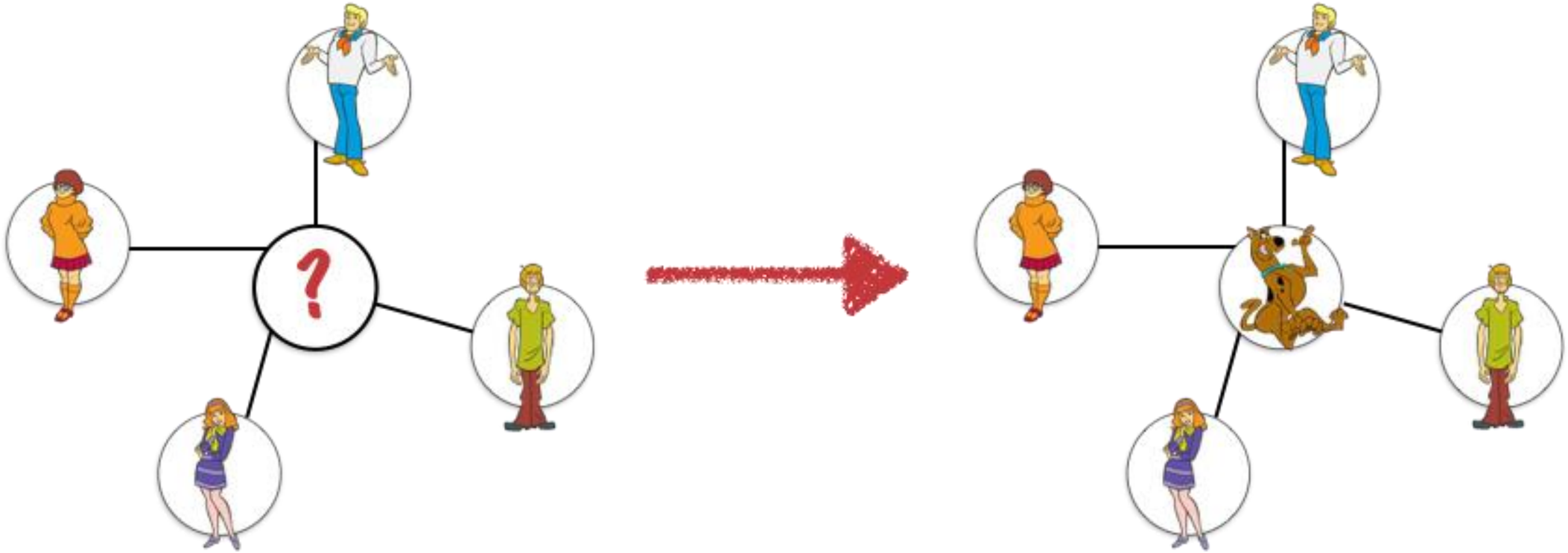


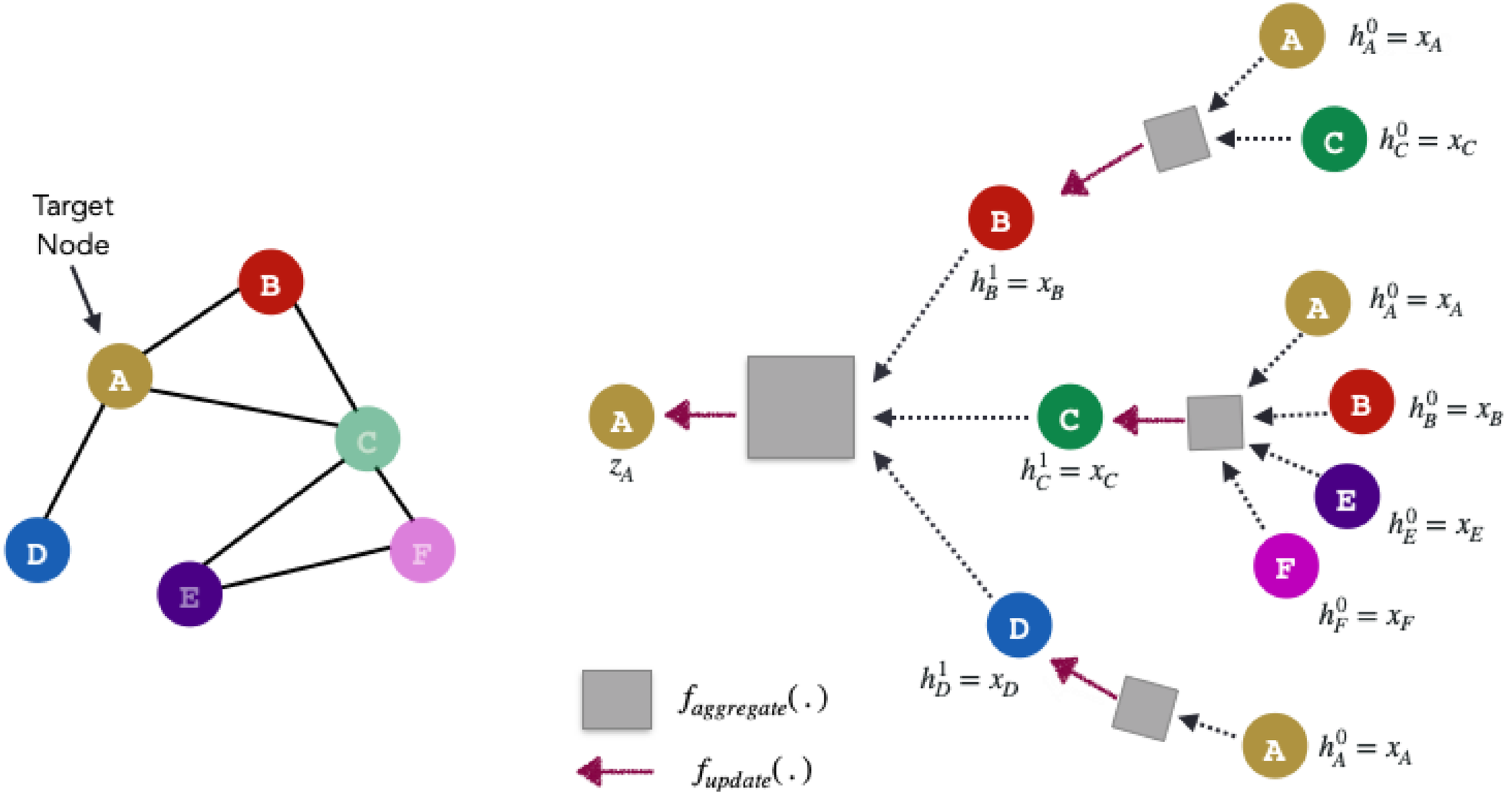
The update rule is now formed as follows:

$$h_i^{(l)} = \sigma(\sum_{i \in N_j} a_{ij} W h_j)$$

3. GraphSAGE

You are known by the company you keep

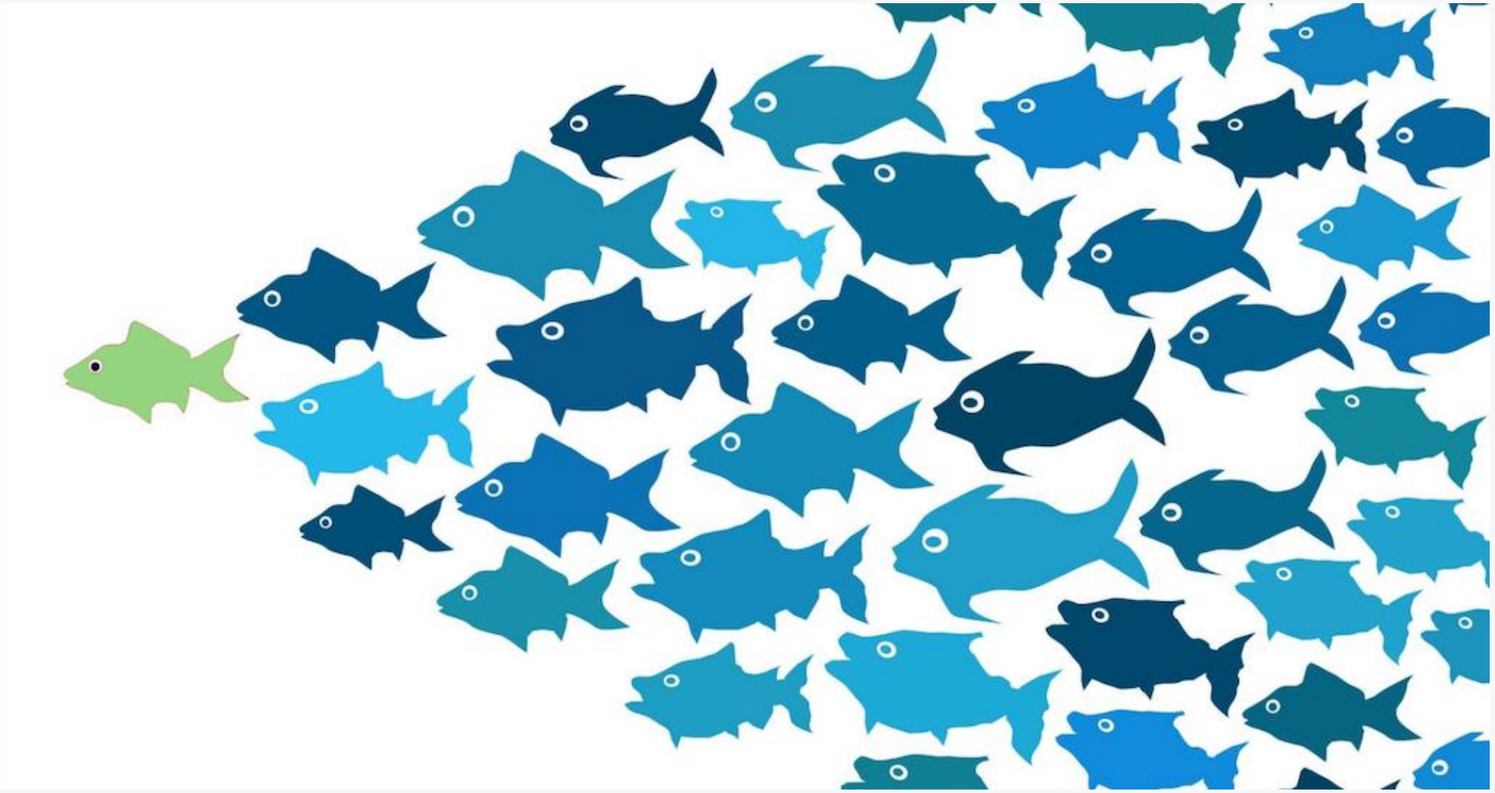



Target
Node
A
B
C
D
E
F
$h_A^0 = x_A$
$h_C^0 = x_C$
$h_B^1 = x_B$
$h_A^0 = x_A$
$h_B^0 = x_B$
$h_E^0 = x_E$
$h_F^0 = x_F$
z_A
$h_C^1 = x_C$
$h_D^1 = x_D$
$h_A^0 = x_A$
$f_{aggregate}(.)$
$f_{update}(.)$

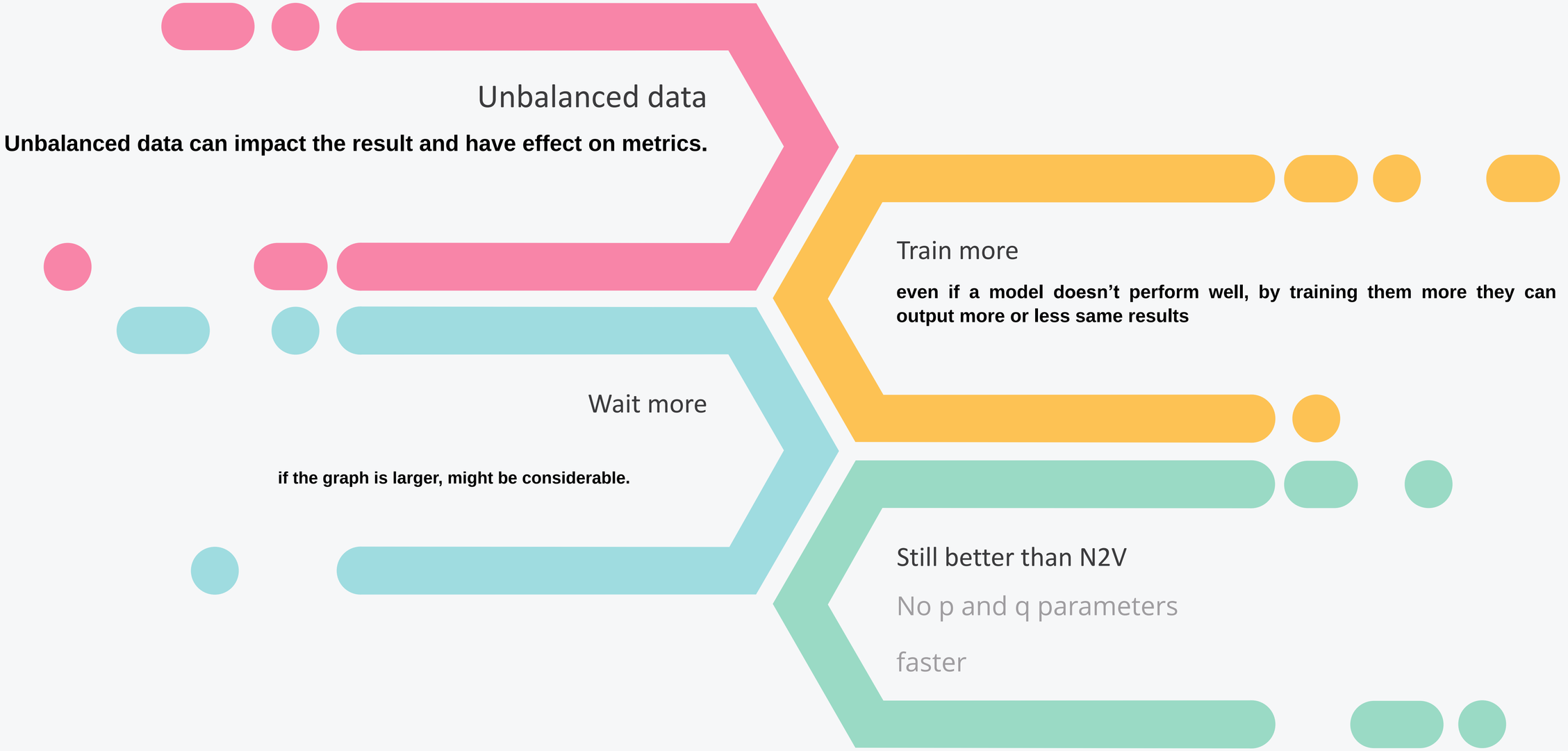

@link https://deve
@package _s
*/
if (! function_exists('incode_starter_setup')) :
/**
* Sets up theme defaults and registers support for various WordPress
* Note that this function is hooked into the after_setup_theme hook, which
* runs before the init hook. The init hook is too late for some features, such
* as indicating support for post thumbnails.
*/
function incode_starter_setup() {
/*
* Make theme available for translation.
* Translations can be filed in the /languages/ directory.
* If you're building a theme based on _s, use a find and replace
* to change 'incode_starter' to the name of your theme in all
st.xml

5. Different Dataset (Unbalanced data)

- ❑ **Is not difficult to predict the results without balancing the data**
- ❑ **Accuracy=1**
- ❑ **other metrics=0**
- ❑ **Unbalancing the graph is crucial, otherwise, no meaningful results will be achieved**
- ❑ **I needed to change the case study since with this data set they only had very few nodes**
- ❑ **The better model= the better Explainer**



6. Conclusion



Thank you