NAOUMI Salmane - KILANI Al Houceine
salmane.naoumi@gmail.com
houcinekila@gmail.com

Lab session # 6
ALTEGRAD 2019

12/26/19

1 Questions :

Task 4

We see here that we managed somehow to group together pages in the same domain (Government, Cultural, Blogs).

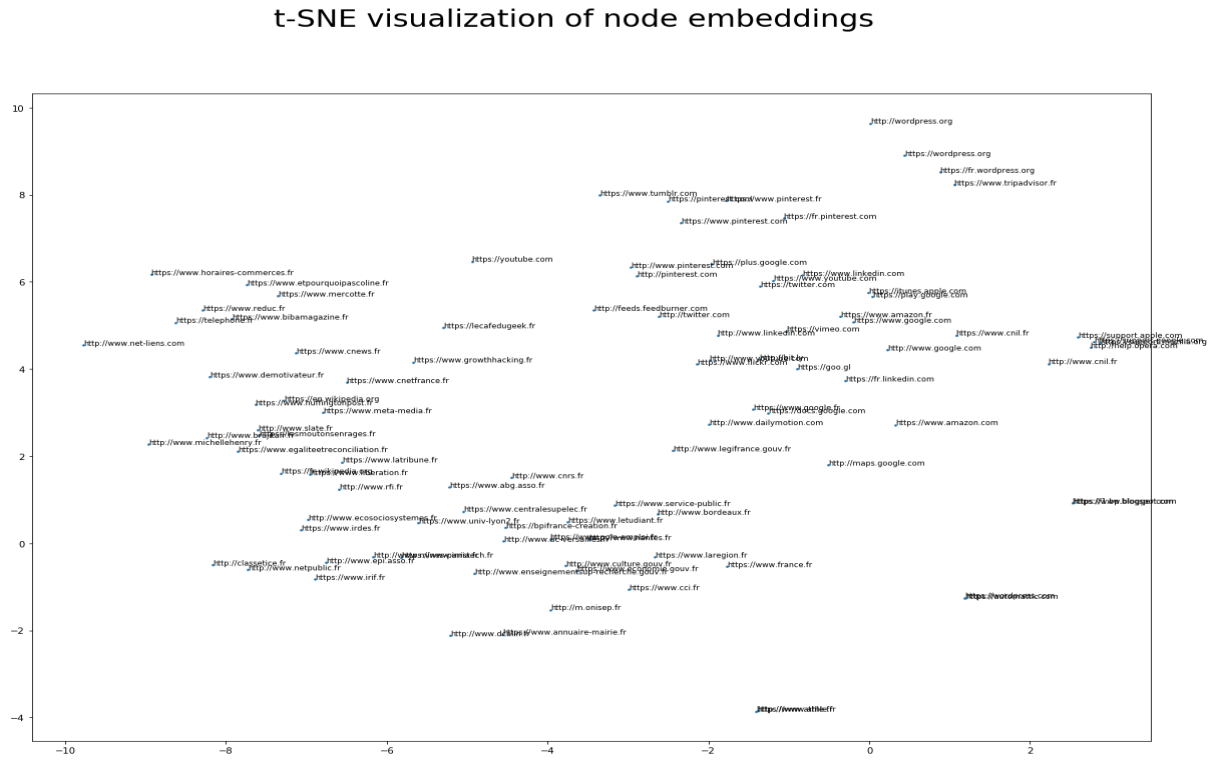

Figure 1: TSNE visualisation of the most recurrent nodes in the walks

Task 5

Both approaches are achieving poor accuracy on the graph. We reach 0.428% classification score with both methods.

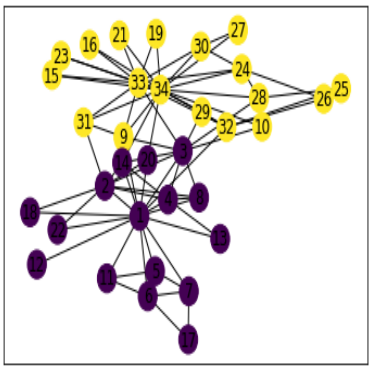

Figure 2: Karate Club graph

1.1 Question 1 :

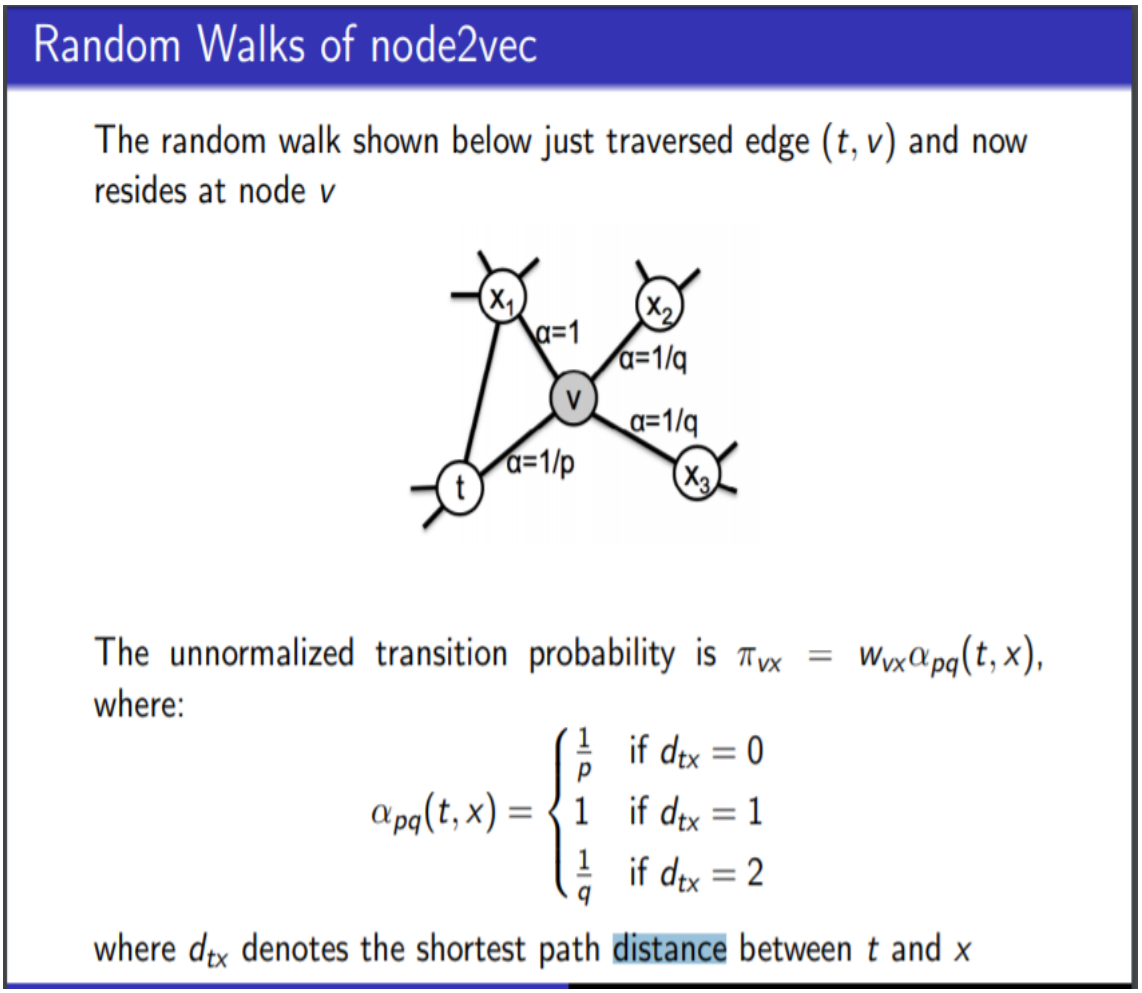

Figure 3: The transition mechanism of Node2Vec (From the course's slides)

We can get inspired by the Node2Vec approach to avoid such behaviours : Instead of randomly sampling (Uniform sampling) of a node in the neighbours of the actual node reached by the walk, we replace this probability by a product between the weight of the edge multiplied by α_{pq} (Figure 3).

Especially, we set p parameter to be high (This the probability of coming back the the starting node is low).

1.2 Question 2 :

A natural tranition from node classification to graph classification would be the following :

- We choose a kernel ϕ (Short parth kernel, graphlets,...)
- a_{ij} the element in the position (i, j) of the A matrix in this case will be $\phi(i)^T\phi(j)$: it measures a similarity between two entire graphs.
- The X matrix will be a matrix which the i-th row corresponds to the vector $\phi(i)$

After that everything remains the same and the GNN will be trainable end-to-end just by changing its inputs. But this kind of adaptation does not take into consideration the information contained in the nodes of the graph if it has any. This is why we think that maybe a concatenation with the nodes representation (obtained by DeepWalk for instance) may give additional information on the graph class (like in the equation 5 of [1])

1.3 Question 3 :

Using a Graph neural network to classify nodes of the Karate Club Dataset, we are able to achieve 100% accuracy on both train and test sets. This approach is outperforming the DeepWalk + logistic regression

method. A reason for that is their effectiveness for learning representations of the graph. we can say that both previous methods generate sub-optimal node representations, DeepWalk for example is considered as an unsupervised learning of embeddings, and thus fail to compete with the GNN.

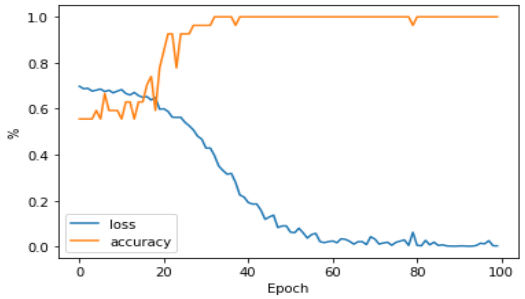

Figure 4: Evolution of training loss and accuracy

1.4 Question 4 :

A change of the initialization of the nodes features showed a huge decrease in the performance of the model (An accuracy of 70%). In fact, GNNs kind of follows a neighborhood aggregation scheme, where the learned representation of a node is computed by recursively aggregating and transforming representations of its neighbouring nodes. Therefore, initializing nodes with the same representation is misleading the model of leaning effective representation vectors.

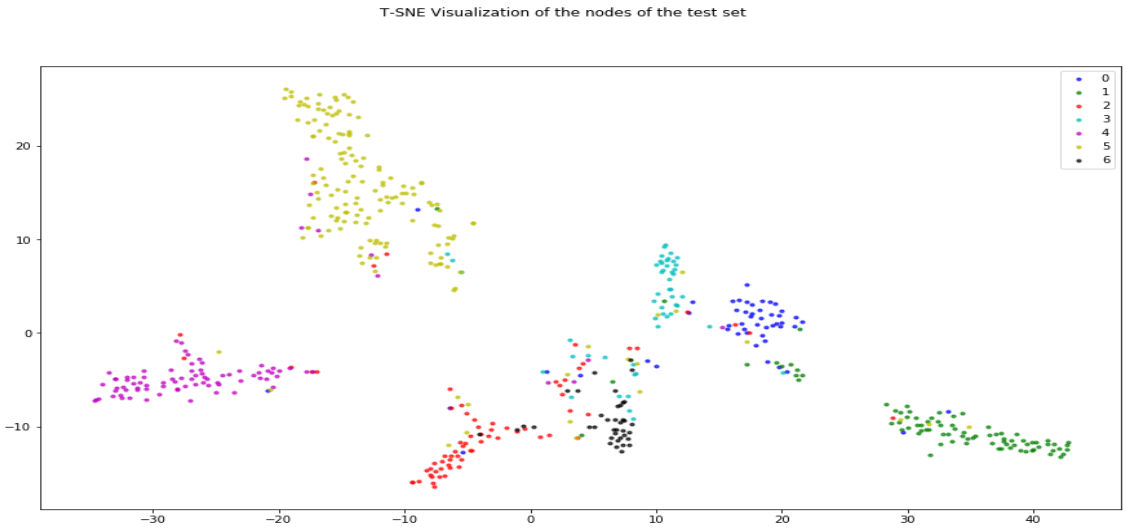

Figure 5: TSNE visualisation of the graphs embeddings

References

[1] T.Chen, ”Are Powerful Graph Neural Nets Necessary?
A Dissection on Graph Classification”, https://arxiv.org/pdf/1905.04579v2.pdf