1 Question 1

So when we do deepwalk, more we add context (here nodes of the neighbourhood) more our embeddings are going to be precise. In the case of K2, each node has exactly one neighbour. Hence the connected component are going to be exactly represented in the random walks. Now when using skip gram it's supposed to learn embeddings with their local neighbourhood, hence here the embedding should represent exactly the connection between 2 nodes. More over two nodes not in the same connected component are never going to be found in the same random walk, hence the skip gram is not going to capture any similarity between those two nodes. All in all, we see that two nodes in the same connected component will learn their embeddings together, hence we can think that the cosine similarity is going to be be near 1 or equal to one in this case. And since on the other hand, two nodes not in the same connected component are embedded knowing nothing from each other, we expect them to have a cosine similarity close to 0. It's not going to be negative since it would show a link between the two nodes.

2 Comparison Deepwalk and Spectral embeddings

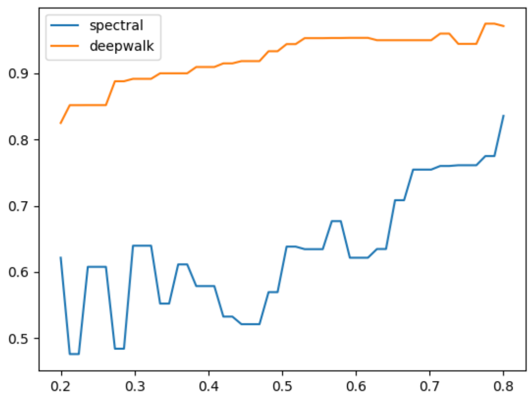

Figure 1: DeepWalk and Sprectal Embeddings classification accuracy

So this graph is the accuracy of both deepwalk and spectral embeddings with respect to the ratio of the training set. To reduce variance, we did the accuracy as a mean of 20 simulation. So, as we can see, it's always better to use deepwalk than the spectral embeddings. Moreover when the training set is really low, the loss in preformance is very little for deepwalk compared to the one for spectral.

3 Question 2

So, in [1], they state the the time complexity for Deepwalk is $O(\gamma|V|tw(d+dlog|V|))$ where V is the vocabulary, γ is the number of random walks, t is the walk length, w the window size and d the representation size. On the other hand, in [3] the complexity proposed is $O(|V|)$, this can be understood that t,w,d and γ are far smaller than $|V|$, hence they can be neglected in the time complexity. All in all we are going to keep the first one that is more precise and more logical when one tries to compute the complexity by oneself. One can note that there is a sum in the complexity since depending on the size of the vocabulary the preponderant term is either gonna be d or $d * log|V|$
For the Spectral Embeddings, in [2] they give us the complexity for the computing of the laplacian matrix and it's eigenvectors which is $O(|V|^3)$ and it's the only thing we do to have the spectral embeddings.
So as we can see the compute price of the spectral embeddings are way higher than for DeepWalk and the performance is worst. Hence it's better overall to use DeepWalk.

4 Question 3

Si if we would to be working with no self loops, here it's what would happen in a single layer of the GNN : So $\hat{A} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$ still has 0 as diagonal coefficient. Let's Z represent the ouput of the previous layer, and W be the weights of the actual layer.
So, when we compute $\hat{A}ZW$, if we look how a row is updated for instace, let's say the third row , which is going to have a 0 at its third coefficient, we have the fact that the new third row, is never going to take in account the last feature of the third row feature vector :
Let ZW = B for simplicity, We then have for the third column $(AB)_{3j} = \sum_h a_{3h} b_{hj} = \sum_{h,h!=3} a_{3h} b_{hj}$ because $a_{33} = 0$. And it's the same for each row of the new feature matrix. Hence, the aggregated messages do not factor in node i's own feature vector. To counterbalance this, we add self nodes that permits us to aggregates feature vectors of its neighbors and its own feature vector to compute its new feature vector.

5 Question 4

Let $\hat{A}$ be defined as in the lab sheet. Hence we have :

$$\hat{a_{i,j}} = \sum_l^n \sum_h^n \tilde{d_{ih}}^{-\frac{1}{2}} \tilde{a_{hl}} \tilde{d_{lj}}^{-\frac{1}{2}}$$

But we know that $\tilde{D}$ is a diagonal matrix, hence it becomes :

$$\hat{a_{i,j}} = \frac{\tilde{a_{ij}}}{\sqrt{\tilde{d_{ii}} \tilde{d_{jj}}}}$$

So now we can start to compute Z^0 and Z^1 using the two weight matrices W^0 and W^1 :

$$W^0 = \begin{bmatrix} 0.5 & -0.2 \end{bmatrix}, W^1 = \begin{bmatrix} 0.3 & -0.4 & 0.8 & 0.5 \\ -1.1 & 0.6 & -0.1 & 0.7 \end{bmatrix}$$

Star Graph :

Using what we just showed we directly have $\hat{A} = \begin{bmatrix} \frac{1}{4} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ \frac{1}{2\sqrt{2}} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2\sqrt{2}} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2\sqrt{2}} & 0 & 0 & \frac{1}{2} \end{bmatrix}$

Then :

$$\hat{A}X = \begin{bmatrix} \frac{1}{4} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ \frac{1}{2\sqrt{2}} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2\sqrt{2}} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2\sqrt{2}} & 0 & 0 & \frac{1}{2} \end{bmatrix} * \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1+3\sqrt{2}}{4} \\ \frac{2+\sqrt{2}}{4} \\ \frac{2+\sqrt{2}}{4} \\ \frac{2+\sqrt{2}}{4} \end{bmatrix}$$

Then :

$$\hat{A}XW^0 = \begin{bmatrix} \frac{1+3\sqrt{2}}{4} \\ \frac{2+\sqrt{2}}{4} \\ \frac{2+\sqrt{2}}{4} \\ \frac{2+\sqrt{2}}{4} \end{bmatrix} * \begin{bmatrix} 0.5 & -0.2 \end{bmatrix} = \begin{bmatrix} \frac{1+3\sqrt{2}}{8} & -\frac{1+3\sqrt{2}}{20} \\ \frac{2+\sqrt{2}}{8} & -\frac{2+\sqrt{2}}{20} \\ \frac{2+\sqrt{2}}{8} & -\frac{2+\sqrt{2}}{20} \\ \frac{2+\sqrt{2}}{8} & -\frac{2+\sqrt{2}}{20} \end{bmatrix}$$

And then we apply the ReLU that set the right column to zero.

$$Z^0 = \begin{bmatrix} \frac{1+3\sqrt{2}}{8} & 0 \\ \frac{2+\sqrt{2}}{8} & 0 \\ \frac{2+\sqrt{2}}{8} & 0 \\ \frac{2+\sqrt{2}}{8} & 0 \end{bmatrix}$$

We can now go to the second layer :

$$\hat{A}Z^0 = \begin{bmatrix} \frac{1}{4} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ \frac{1}{2\sqrt{2}} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2\sqrt{2}} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2\sqrt{2}} & 0 & 0 & \frac{1}{2} \end{bmatrix} * \begin{bmatrix} \frac{1+3\sqrt{2}}{8} & 0 \\ \frac{2+\sqrt{2}}{8} & 0 \\ \frac{2+\sqrt{2}}{8} & 0 \\ \frac{2+\sqrt{2}}{8} & 0 \end{bmatrix} = \begin{bmatrix} \frac{7+9\sqrt{2}}{32} & 0 \\ \frac{10+3\sqrt{2}}{32} & 0 \\ \frac{10+3\sqrt{2}}{32} & 0 \\ \frac{10+3\sqrt{2}}{32} & 0 \end{bmatrix}$$

Then :

$$\hat{A}Z^0W^1 = \begin{bmatrix} \frac{7+9\sqrt{2}}{32} & 0 \\ \frac{10+3\sqrt{2}}{32} & 0 \\ \frac{10+3\sqrt{2}}{32} & 0 \\ \frac{10+3\sqrt{2}}{32} & 0 \end{bmatrix} * \begin{bmatrix} 0.3 & -0.4 & 0.8 & 0.5 \\ -1.1 & 0.6 & -0.1 & 0.7 \end{bmatrix} = \begin{bmatrix} 0.3\frac{7+9\sqrt{2}}{32} & -\frac{7+9\sqrt{2}}{80} & \frac{7+9\sqrt{2}}{40} & \frac{7+9\sqrt{2}}{64} \\ 0.3\frac{10+3\sqrt{2}}{32} & -\frac{10+3\sqrt{2}}{80} & \frac{10+3\sqrt{2}}{40} & \frac{10+3\sqrt{2}}{64} \\ 0.3\frac{10+3\sqrt{2}}{32} & -\frac{10+3\sqrt{2}}{80} & \frac{10+3\sqrt{2}}{40} & \frac{10+3\sqrt{2}}{64} \\ 0.3\frac{10+3\sqrt{2}}{32} & -\frac{10+3\sqrt{2}}{80} & \frac{10+3\sqrt{2}}{40} & \frac{10+3\sqrt{2}}{64} \end{bmatrix}$$

And then we apply ReLU and we have :

$$Z^1 = \begin{bmatrix} 0.3\frac{7+9\sqrt{2}}{32} & 0 & \frac{7+9\sqrt{2}}{40} & \frac{7+9\sqrt{2}}{64} \\ 0.3\frac{10+3\sqrt{2}}{32} & 0 & \frac{10+3\sqrt{2}}{40} & \frac{10+3\sqrt{2}}{64} \\ 0.3\frac{10+3\sqrt{2}}{32} & 0 & \frac{10+3\sqrt{2}}{40} & \frac{10+3\sqrt{2}}{64} \\ 0.3\frac{10+3\sqrt{2}}{32} & 0 & \frac{10+3\sqrt{2}}{40} & \frac{10+3\sqrt{2}}{64} \end{bmatrix}$$

Cycle Graph :

Using what we just showed we directly have $\hat{A} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$

Then :

$$\hat{A}X = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix} * \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Then :

$$\hat{A}XW^0 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} * \begin{bmatrix} 0.5 & -0.2 \end{bmatrix} = \begin{bmatrix} 0.5 & -0.2 \\ 0.5 & -0.2 \\ 0.5 & -0.2 \\ 0.5 & -0.2 \end{bmatrix}$$

Now we apply ReLU and we have :

$$Z^0 = \begin{bmatrix} 0.5 & 0 \\ 0.5 & 0 \\ 0.5 & 0 \\ 0.5 & 0 \end{bmatrix}$$

We can now go to the second layer :

$$\hat{A}Z^0 = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix} * \begin{bmatrix} 0.5 & 0 \\ 0.5 & 0 \\ 0.5 & 0 \\ 0.5 & 0 \end{bmatrix} = \begin{bmatrix} 0.5 & 0 \\ 0.5 & 0 \\ 0.5 & 0 \\ 0.5 & 0 \end{bmatrix}$$

Then :

$$\hat{A}Z^0W^1 = \begin{bmatrix} 0.5 & 0 \\ 0.5 & 0 \\ 0.5 & 0 \\ 0.5 & 0 \end{bmatrix} * \begin{bmatrix} 0.3 & -0.4 & 0.8 & 0.5 \\ -1.1 & 0.6 & -0.1 & 0.7 \end{bmatrix} = \begin{bmatrix} 0.15 & -0.2 & 0.4 & 0.25 \\ 0.15 & -0.2 & 0.4 & 0.25 \\ 0.15 & -0.2 & 0.4 & 0.25 \\ 0.15 & -0.2 & 0.4 & 0.25 \end{bmatrix}$$

And then we apply ReLU and we have :

$$Z^1 = \begin{bmatrix} 0.15 & 0 & 0.4 & 0.25 \\ 0.15 & 0 & 0.4 & 0.25 \\ 0.15 & 0 & 0.4 & 0.25 \\ 0.15 & 0 & 0.4 & 0.25 \end{bmatrix}$$

As we can see for both of the Z^1 we calculated, we have a lot of coefficient in the matrix that are the same, for instance with the cycle graph all the rows are the same, and for star graph, if you let go the first row, all the other rows are also the same. Hence as we see nodes with similar structure (ie neigbourhood) have the same structure in our architecture and we have this a lot in our two graph type that are very regular. One can counterbalance this by not initialising X as a vector with all features equal.

References

[1] Yifan Hu Steven Skiena Haochen Chen, Bryan Perozzi. Harp: Hierarchical representation learning for networks. 2017.

[2] Masahiro Takatsuka Mashaan Alshammari. Approximate spectral clustering with eigenvector selection and self-tuned k. 2023.

[3] Tiago Pimentel Rafael Castro Adriano Veloso Nivio Ziviani. Efficient estimation of node representations in large graphs using linear contexts. 2019.