Question 1

The number of columns of Z represent the compressed information of the nodes of the graph. As we have $\hat{A}_{i,j} = \frac{1}{1+e^{-z_i z_j}}$, the probability of the presence of an edge between i and j increases when z_i and z_j are colinear. If there are more columns, there are more elements in z_i and z_j and therefore more degrees of freedom for the encoder to learn how to represent edges with the rows of Z. If there are more columns than nodes, than the encoder does not learn how to represent nodes but rather how to represent edges with a higher dimension matrix. When running the autoencoder with a nigger $nhidden2$, we observe that the loss function decreases but that the PCA plot of the embeddings does not seperate the points very well. This corresponds to the phenomenon described. The autoencoder "overfits" the edges and forgets the nodes' caracteristics.

Question 2

A possible initialization of X where each node can be uniquely identified is to take the identity matrix.

Question 3

There is $\binom{n}{2}$ possible edges in Erdos-Rényi random graph and each edge is present with probability p independently of other edges. Therefore by linearity of the Expected Value, the expected number of edges on a random graph is $\binom{n}{2}p$. The variance of the number of edges is $\binom{n}{2}p(1-p)$. For $p = 0.2$ and $n = 15$, we have an expectation of 21 and a variance of 16.8. For $p = 0.4$ and $n = 15$ we have an expectation of 42 and a variance of 25.2.

Question 4

Sum : For the Sum readout, we have $\begin{pmatrix} Z_{G_1} \\ Z_{G_2} \\ Z_{G_3} \end{pmatrix} = \begin{pmatrix} 0.7 & -0.49 & 2.39 \\ 3 & 0.4 & 2 \\ 1.5 & 0.2 & 1 \end{pmatrix}$

Mean : For the Mean readout, we have $\begin{pmatrix} Z_{G_1} \\ Z_{G_2} \\ Z_{G_3} \end{pmatrix} = \begin{pmatrix} 0.23 & -0.16 & 0.8 \\ 0.75 & 0.1 & 0.5 \\ 0.75 & 0.1 & 0.5 \end{pmatrix}$

Max : For the Max readout, we have $\begin{pmatrix} Z_{G_1} \\ Z_{G_2} \\ Z_{G_3} \end{pmatrix} = \begin{pmatrix} 0.89 & 0.34 & 1.31 \\ 0.89 & 0.34 & 1.31 \\ 0.89 & 0.34 & 1.31 \end{pmatrix}$

The Mean readout gives the same embedding for G_2 and G_3, the Max readout gives the same embedding for all graphs, whereas the sum readout gives different embeddings for each graph. Therefore the sum readout is able to distinguish the three graphs the best.