1 Question 1

Let n be the number of nodes in the graph and p be the edge probability. Let $X_{ij} = X_{ji} = 1$ indicate that an edge exists between nodes i and j (we always have $X_{ii} = 0$). Clearly $\mathbb{E}[X_{ij}] = p$. Let's consider a node i, it can be connected to the remaining $n-1$ nodes independently, with probability p each time.

$$\begin{aligned}
\mathbb{E}[\deg(v)] &= \mathbb{E}\left[\sum_{j=1}^{n} X_{ij}\right] \\
&= \sum_{j=1}^{n} \mathbb{E}[X_{ij}] && \text{by independance.} \\
&= (n-1)p && \text{since } X_{ii} = 0.
\end{aligned}$$

2 Question 2

Using a sum or a mean as a readout function is quite intuitive : each node has the same contribution to the final value. Can we use a fully connected layer instead ?

- For one thing, using only a fully connected layer we realize that we cannot go from $\mathbb{R}^{n\times h}$ to $\mathbb{R}^{h}$ (we would have to transpose the features matrix first).
- If we transpose the features matrix first, we can use a fully connected layer to go from $\mathbb{R}^{h\times n}$ to $\mathbb{R}^{h}$. The weights matrix would be of size $n \times 1$, each coefficient being the importance of a given node towards the final value.

So it comes down to whether we want the model to learn node-wise importance. Since such an operation is not invariant to permutations of the nodes, it is not desirable. Furthermore, n may not be known in advance when we are using mini-batches. That's why we use a sum or a mean as a readout function.

3 Question 3

Aggregation \ Readout	Mean	Sum
Mean	All graphs have the same feature vector.	Different graphs have different features with increasing amplitudes with the size of the graph.
Sum	All graphs have the same feature vector. Their amplitudes are greater than with mean aggregation.	Different graphs have different features with increasing amplitudes with the size of the graph. Their amplitudes are greater than with mean aggregation.

These results are quite surprising, we would expect the sum readout and the mean readout to have the same expressiveness (ability to distinguish between graphs). However, the sum readout seems to be more expressive. We can explain this by the fact that the mean readout will normalize the features by the number of nodes. In this case where all features are identical, the mean readout will always give the same result even when the number of nodes varies. The sum readout will give different results depending on the number of nodes.

4 Question 4

Taking into account what we discovered in the previous question, a first condition for two graphs to be indistinguishable with the sum readout is that they have the same number of nodes. Similarly to how $C_3 \cup C_3$ and C_6 are indistinguishable with the sum readout, we have that $C_4 \cup C_4$ and C_8 to also be indistinguishable, and for any $n \geq 3$, $C_n \cup C_n$ and C_{2n} are indistinguishable.