Homework 2

1) Sample space

We are operating in graph space, $\mathcal{G}_n = \{\mathcal{V}, \mathcal{E}, \mathcal{Y}\}$ where
$\mathcal{V} = \{v_1, v_2, \dots, v_n\}$ is a set of vertices,
$\mathcal{E} = \{e_{11}, \dots, e_{nn}\}$ is a set of potential edges and
$\mathcal{Y} = \{0,1\}^n$ (each vertex has a label 0 or 1)

2) Model

We used the Stochastic Block Model, $SBM_n^k(\vec{p}, \vec{\beta})$ where $\vec{p} \, \epsilon \, \Delta_2 \; and \; \vec{\beta} \, \epsilon \, (0,1)^{2x2}$. $k = 2$ since all vertices will be either 0 or 1.

3) Action space

$\mathcal{A} = \{y \, \epsilon \, \{0,1\}^n\}$
All possible cluster assignments for each vertex i.

4) Decision Rule class

The clustering method used, e.g. kmeans.

5) Loss function

$\ell: \mathcal{G}_n x \mathcal{A} \rightarrow \mathbb{R}_+$
The loss could be measured as $\ell = \sum_{i=1}^{n} \mathbb{I}\{\widehat{y_i} = y_i\}$
or a better way to measure the loss would be using the adjusted rand index (ARI) where
$ARI = \frac{Index - E[I]}{Max(I) - E[I]}$

6) Risk function

$\mathcal{R}: \mathcal{P} x \ell \rightarrow \mathbb{R}_+$, e.g. $E_p[\ell]$ (the expected value of the loss function)