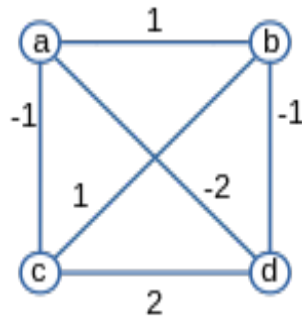


**MACHINE LEARNING 1 WS2019/20**  
**4. EXERCISE**

- Clustering -

**problem 1.** Write down all possible clusterings and their cost for the following graph:



**problem 2.** You want to cluster  $N$  elements into  $C$  clusters. Let the size of the clusters be equal to  $K$  and  $N = C \cdot K$ . How many internal and external edges does a fully connected graph with all elements have? (e.g.  $N=1000$ ,  $C=10$ )

**problem 3.** Download the [MNIST data](#). For the first 1000 (e.g.) images calculate the cost matrix as the negative  $L^2$ -norm between every image pair. Add a constant to allow a choosable percentage of cost values to be positive. Consider the fully connected graph of all images. Use the Kernighan-Lin algorithm to cluster that graph and give an interpretation of the result. You may use this [modified Kernighan-Lin code](#). In addition change it to have a fixed number of clusters. Compare the results.

For reading and writing images additional libraries are not necessary in this case but may be used - e.g. OpenCV, scikit-image. Here is a [starting code](#) to read the data and calculate the cost matrix.