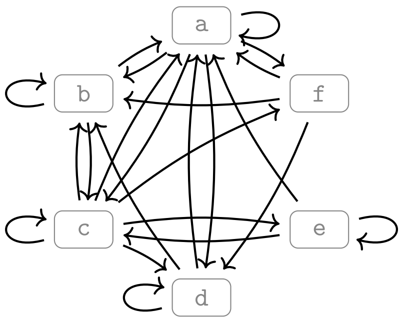
Please carefully read and follow the instructions regarding course work submissions. Failing to meet the requirements might lead to penalties. https://elearn.uef.fi/mod/page/view.php?id=135658

If you suspect that something is wrong with some exercise question, please contact the lecturer.

If you face persistent issues while working on an exercise, do ask for help, e.g. during a course meeting or by contacting the lecturer via email.

We consider the following directed graph over six nodes, and walks over this network. We assume that *(i)* a walk may start in any node with equal probability, *(ii)* at each step, the next node is chosen uniformly at random among the neighbors of the current node and *(iii)* the walk stops after at least three steps. Two such walks are shown below, to the right of the graph.



e → c → f

a → a → d → b → a

Problem 1 (Entropies).
Consider the random variables R_0, R_1, R_2, and so on, representing the position at the start of the walk, after the first step and after the second step, etc.
a) Compute the entropies $H(R_0)$, $H(R_1)$ and $H(R_2)$.
b) Compute the conditional entropies $H(R_0 \mid R_1)$ and $H(R_1 \mid R_0)$, the joint entropy $H(R_0, R_1)$ and the mutual information $I(R_0; R_1)$, as well as for R_1 and R_2. Show intermediate calculation steps, for obtaining transition probabilities, etc.

Problem 2 (Encoding a graph).
We want to transmit the directed graph through a lossless binary channel. To do so, we need to transmit the list of edges, as a list of ordered pairs of nodes, encoded in binary format.

First we decide to use fixed-length codewords for the nodes.
a) Assign codewords to the nodes and write the message that represents the graph.

Next, we decide to use variable-length codewords for the nodes, determined by means of Huffman coding.
b) Assign codewords to the nodes and write the message that represents the graph.

Finally, assume that we ignore the integer code length constraint.
c) What is the length of the message using an optimal frequency-based code?

Problem 3 (Encoding walks).
We want to transmit walks over the graph. To do so, we need to transmit the ordered list of nodes encountered during the path, encoded in binary format.

First we decide to use fixed-length codewords for the nodes.
a) Assign codewords to the nodes and write the message that represents the two walks above.

If we assume that the receiver knows the structure of the graph, because we transmitted it earlier, we can exploit this information. That is, we still need to specify the starting node, but then we can assume an ordering of the nodes and specify which, among the ordered neighbors, is the next node. For instance, if we are currently in node e and assume the nodes are sorted lexicographically, the neighbors of the current node are, in order, a,

c and e. Then, rather than stating that c is the next node, we can say it is the second neighbor. After the first three steps, we need to account for the possiblity that the walk stops.
b) Design a code following this strategy and write the message that represents the two walks above.

Note that while we consider a small toy example, the benefits of this strategy would be significant for a graph where the degree of the nodes is much smaller than the total number of nodes.