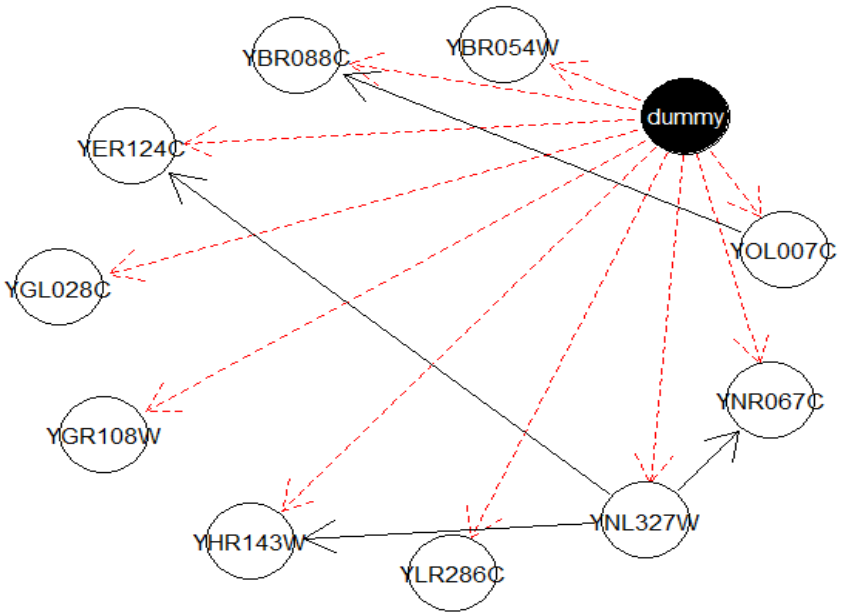
Project 1 - Kacper Domżał 386308

1-2. Run the preprocessing steps of the R-script to filter out the 10 most variant genes. Create a prior structure for the network by hand (see the instructions in the R-script). Regarding the prior structure, suppose that YOL007C has an effect on YBR088C, and that YNL327W has an effect on YER124C, YHR143W and YNR067C. Plot your prior structure and include it in your report.



3. Inspect the local probability distributions either by clicking on the nodes or via the localprob command. What is the output for gene YBR088C?

$YBR088C

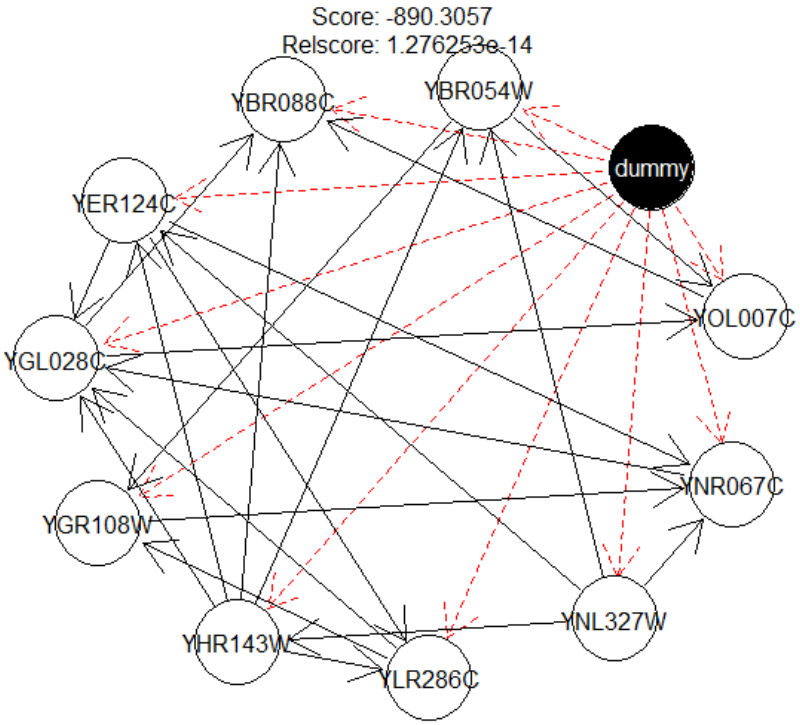
s2	Intercept: YBR088C	YOL007C
0.228734538	0.008498828	0.826678327

4-5. Generate a prior distribution for the parameters of the joint distribution using the jointprior command. Learn the initial Bayesian network. What is its score?

The score is: -1101.865

6. Perform a local search for an optimal network. Plot the network obtained and include it in your report. This optimal network is called BN*. What is the score now?

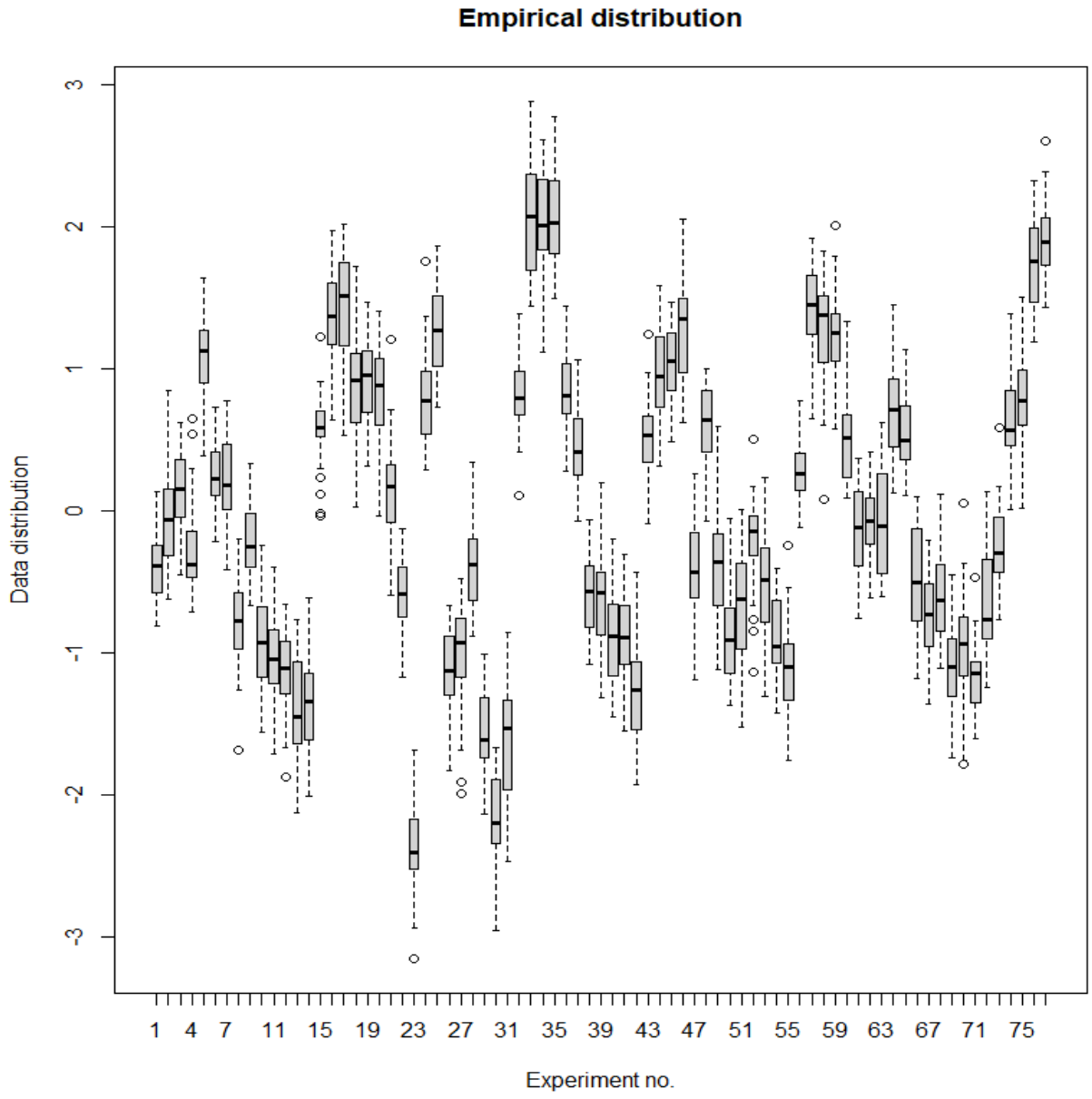
The score is: -890.3057



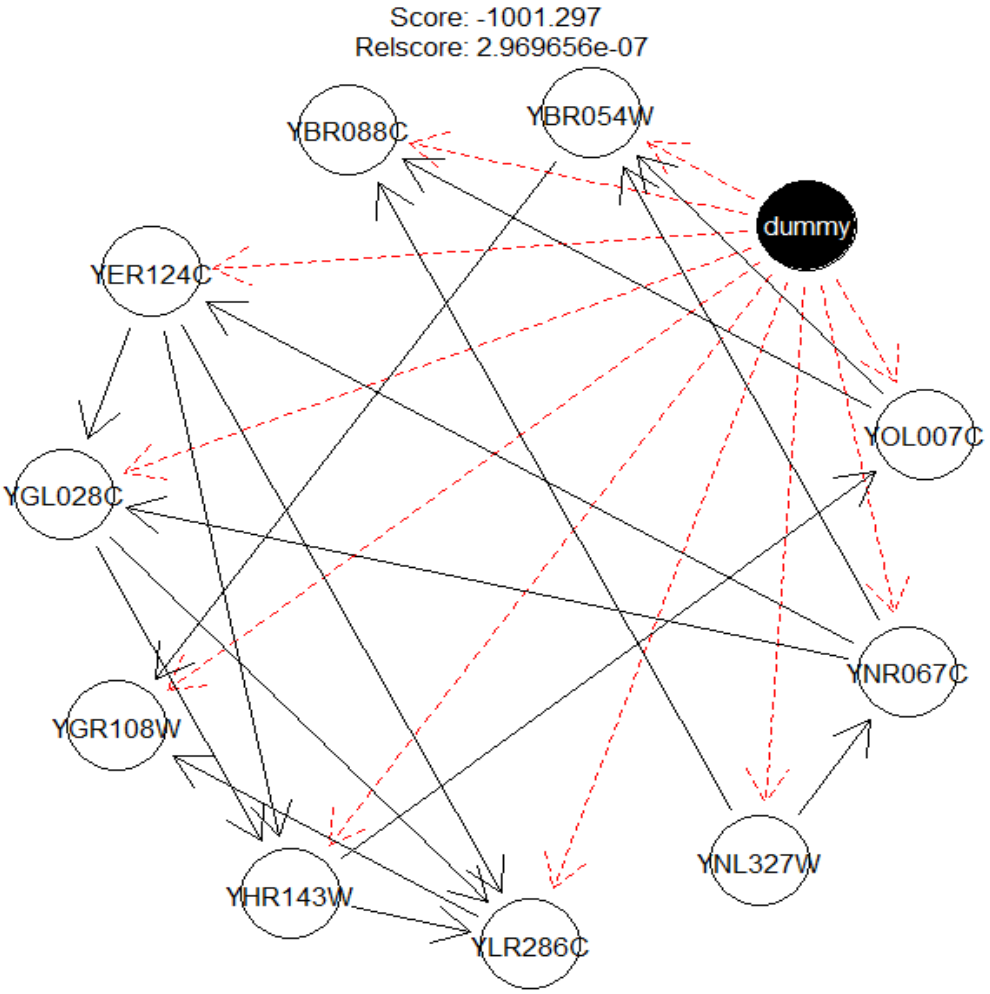
7. For each of the 10 genes calculate the variance sigma i 2 among the experiments and include them in your report.

YBR054W	YBR088C	YER124C	YGL028C	YGR108W	YHR143W	YLR286C	YNL327W	YNR067C	YOL007C	dummy
1.4410599	0.9897524	1.4331767	1.4743219	0.9366332	1.0552875	2.0208231	1.4356611	1.6242167	1.1091768	0.0000000

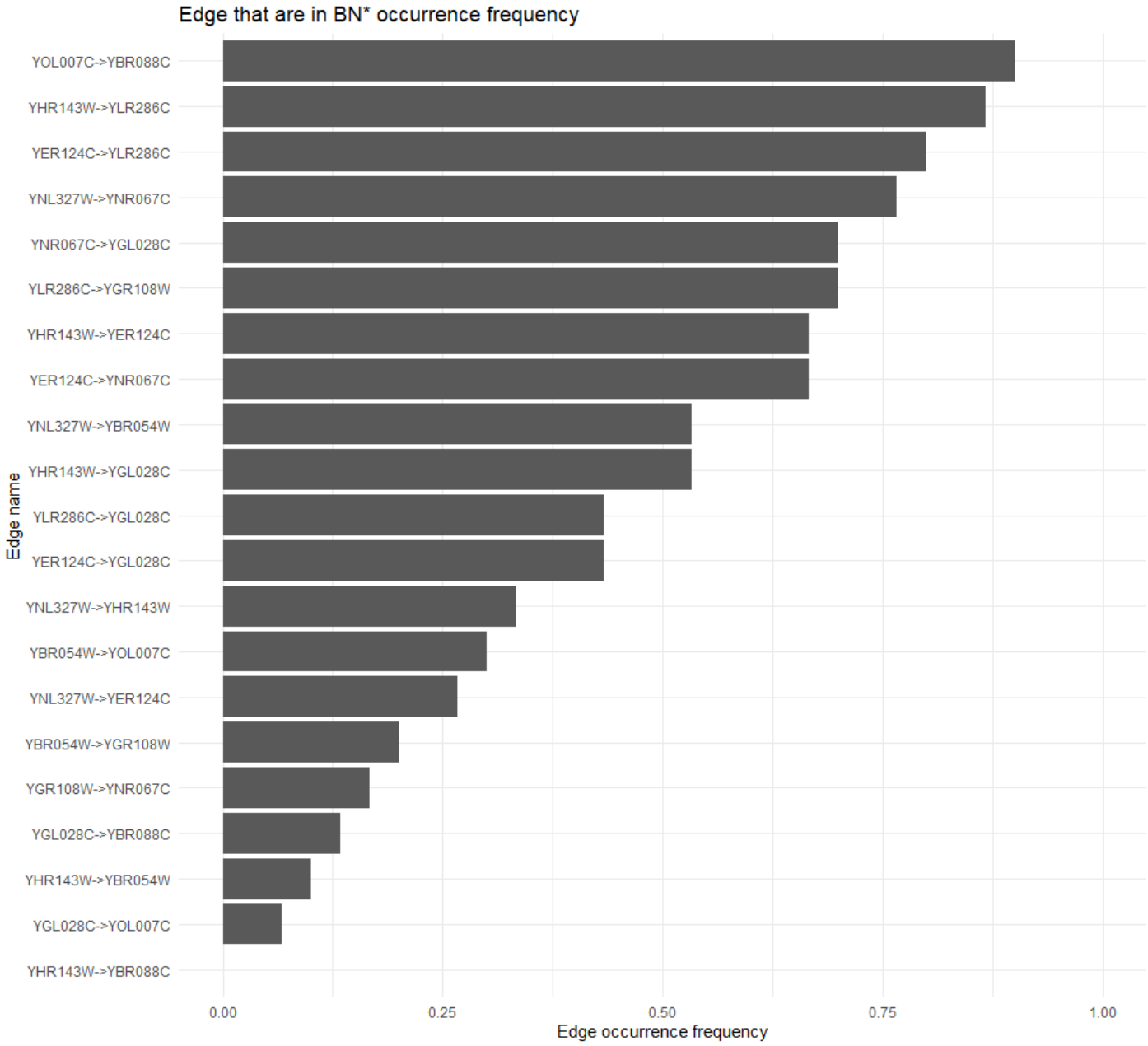
8 -9. Perturb the experimental values of each gene i by adding a noise term distributed as N(0, sigmai 2 10) to each entry in the column corresponding to gene i. Repeat this procedure 30 times for each gene such that you generate 30 perturbed datasets. For gene YHR143W, generate the following box plot: The horizontal axis should display each experiment, i.e., the values 1 to 77. The vertical axis should display one box per experiment.



10. Repeat steps 2-6 (this time ignore the requests for your report) with each of the 30 perturbed datasets. The optimal networks obtained are called PBN1, …, PBN30. Plot the graph corresponding to the network PBN5 and include it in your report.

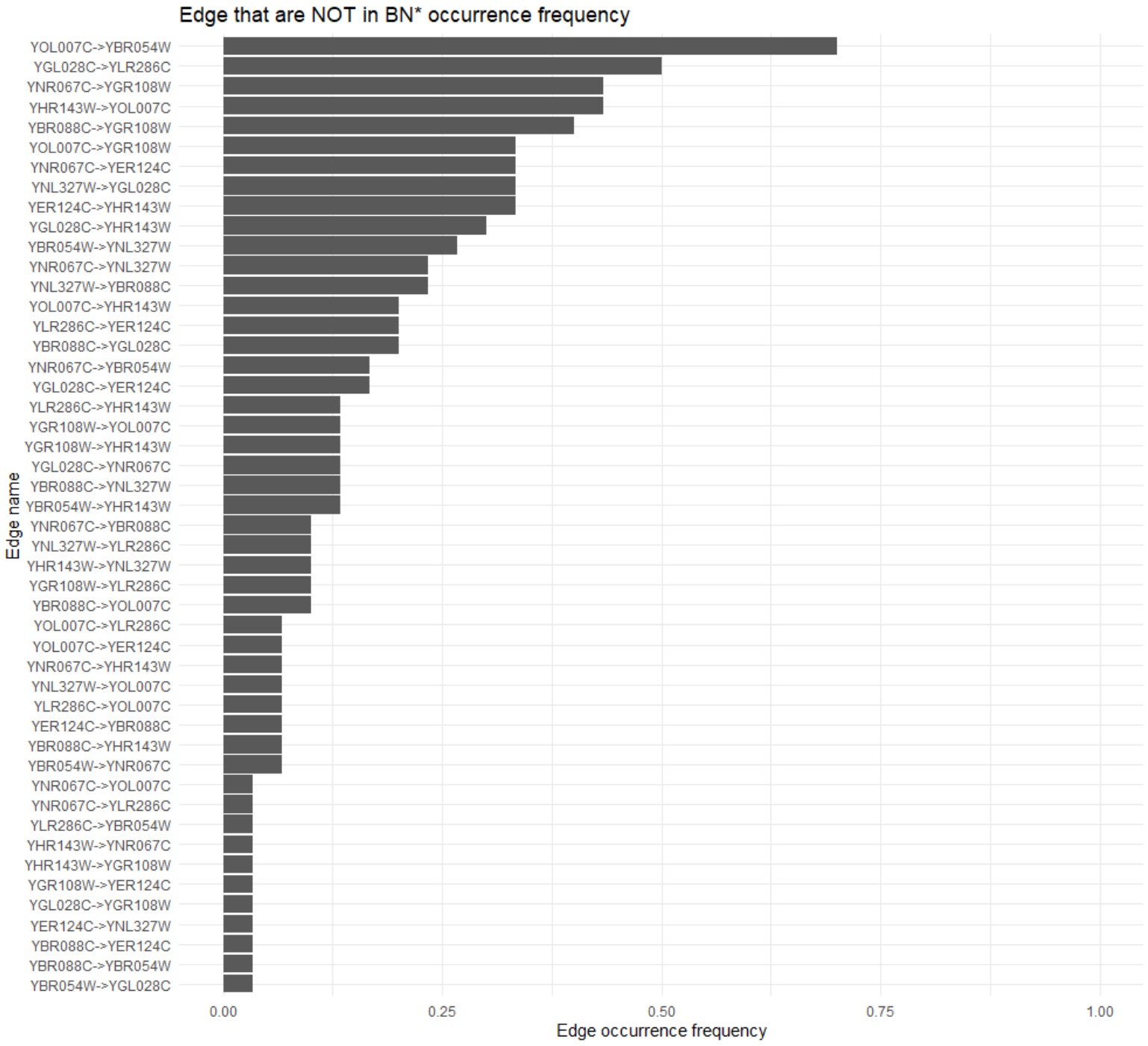


11. For each edge contained in BN*, calculate the relative frequency of appearance among the PBNi. Plot these frequencies and include this in your report. Which edges of your optimal network seem to be spurious?



As you can see edges YHR143W → YBR088C and YGL028C -> YOL007C can be considered spurious, because they appear in none/ very few PBNs. Also being spurious is a relative term so i just chose 2 least frequently appearing edges among all edges.

12. For each edge not contained in BN*, establish its relative frequency among the PBNi. Plot these frequencies and include this in your report. Are there edges that might be missing in your optimal network?



According to the barplot above there are edges that are present in many networks like: YOL007C → YBR054W and YGL028C →YLR286C that might be missing in my optimal network.