AML Exercise 7 – 08.06.17 – Ben Wulf, Lie Hong, Amnon Bleich

Task 1.1

Use the exhaustive search inference algorithm to find the best scoring model.

log posterior (marginal) likelihood $mLL: -239.9944

Error probabilities alpha and beta: 0.15 0.07

network structure regularization parameter $lambda (default: 0): 0

Prior weight $delta for assigning E-genes to virtual S-gene 'null' (default: 1): 1

67 selected E-genes:

--> rel : 30 attached E-genes

--> key : 30 attached E-genes

--> tak : 9 attached E-genes

--> mkk4hep : 28 attached E-genes

--> null : 1 attached E-genes

Task 1.2

To which S genes are the E genes most likely connected to?

AttA rel

CG8177 rel

KrT95D tak

CG6725 mkk4hep

RhoL mkk4hep

CG1141 mkk4hep

puc mkk4hep

CG5775 mkk4hep

wun mkk4hep

CG10076 rel

CG5835 mkk4hep

CG1225 rel

CG5346 mkk4hep

Su(dx) tak

CG8008 rel

CecA1 rel

CecA2 rel

CecB rel

CecC rel

Gli mkk4hep

Dro rel

Mtk rel

EG:95B7.1 mkk4hep

Nhe3 mkk4hep

CG11709 rel

CG12703 tak

CG13780 mkk4hep

CG13117 mkk4hep

CG11066 mkk4hep

CG15900 mkk4hep

CG8046 rel

CG8805 mkk4hep

CG3884 mkk4hep

CG6701 rel

CG12505 rel

CG18372 rel

CG10794 rel

CG15678 rel

CG4859 mkk4hep

CG13893 mkk4hep

CG18214 tak

CG9208 mkk4hep

CG4437 rel

CG6449 mkk4hep

CG14567 rel

CG7629 rel

CG7142 rel

CG7956 tak

CG3348 tak

GH13327 rel

HL1913. rel

LD32282 tak

Rel rel

loco tak

CG7816 rel

CG7778 rel

CG14704 rel

shn rel

CG13503 mkk4hep

Rac2 mkk4hep

Fim mkk4hep

AnnIX mkk4hep

lama tak

CG17723 mkk4hep

CG11798 rel

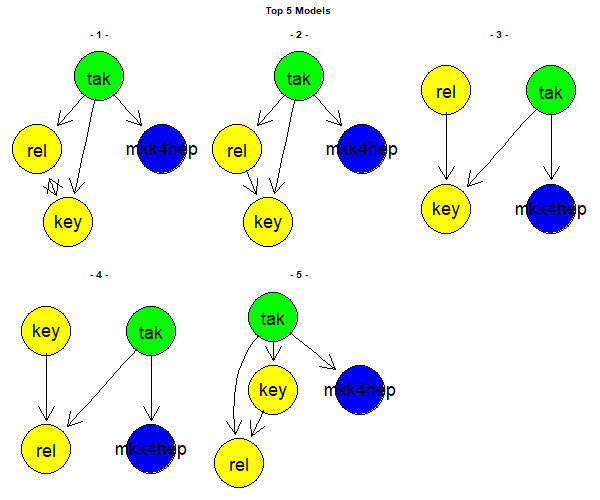
Jra mkk4hep

CG4057 mkk4hep

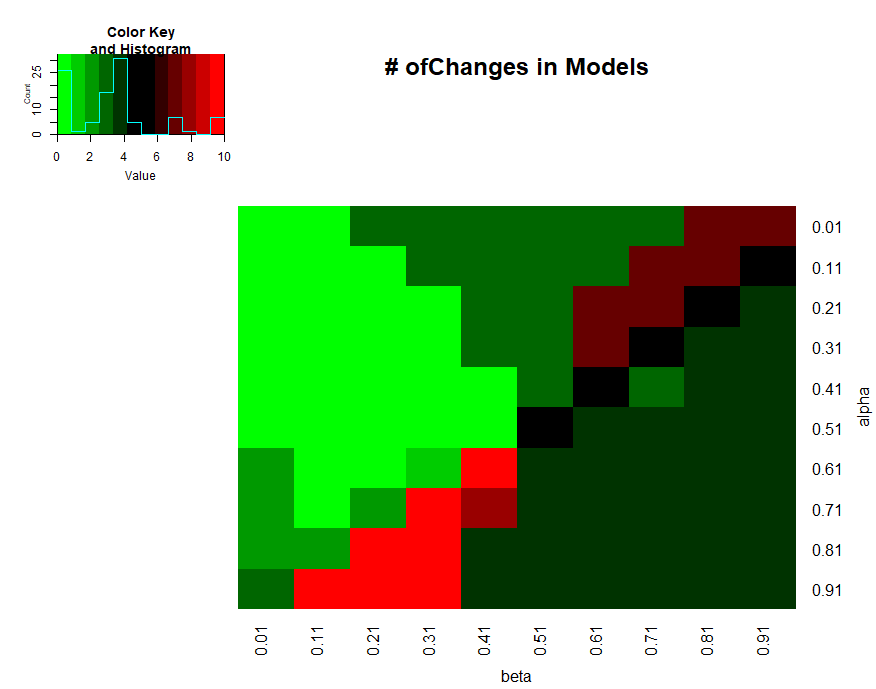
CG11141 null

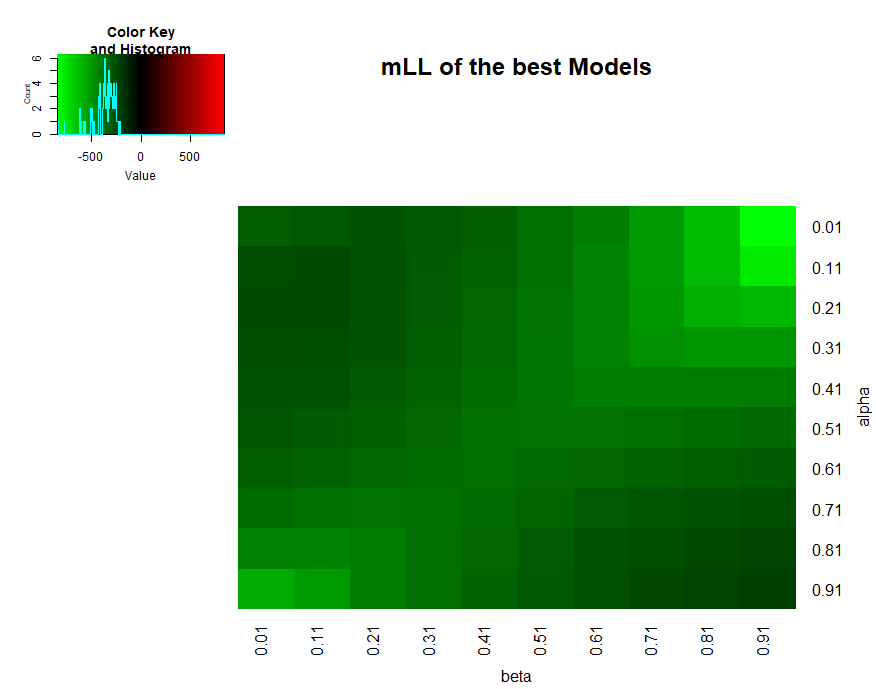
Task 1.3  
How do the top 5 exhaustive models differ?

Model 1,2 and 5 are nearly similar but the double edges between key and rel are in 2 and 5 are only single edges. Model 3,4 are also equal with changed positions of rel and key.



Task 1.4  
Evaluate stability for different choices of alpha, beta.

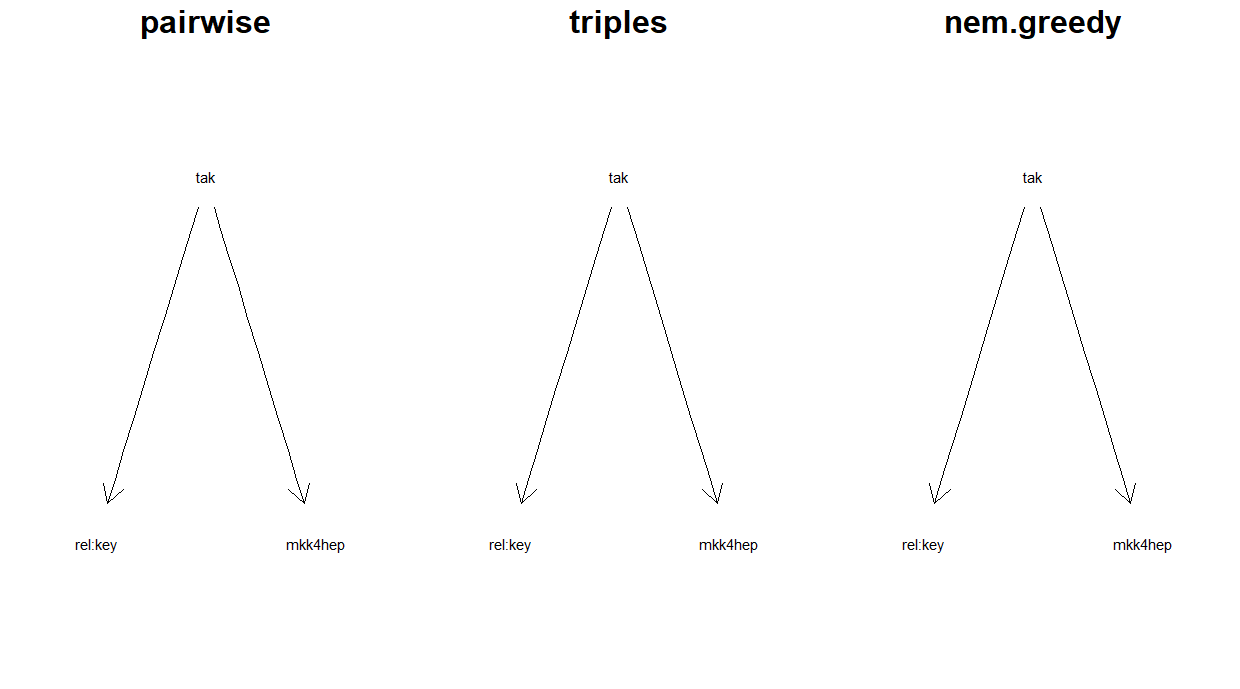
We test different choices for the alpha and beta error and count the differences to the model (alpha and beta: 0.15 0.07)with the calculated errors. The number of changes increase with the difference in the alpha and beta errors. The error rates allow a bigger verity in the model.



In the histogram, there of the mLL are better scores darker. Thus, medium changes in alpha and beta error will lead to a model with a lower probability while big changes in both will lead to the original quality.

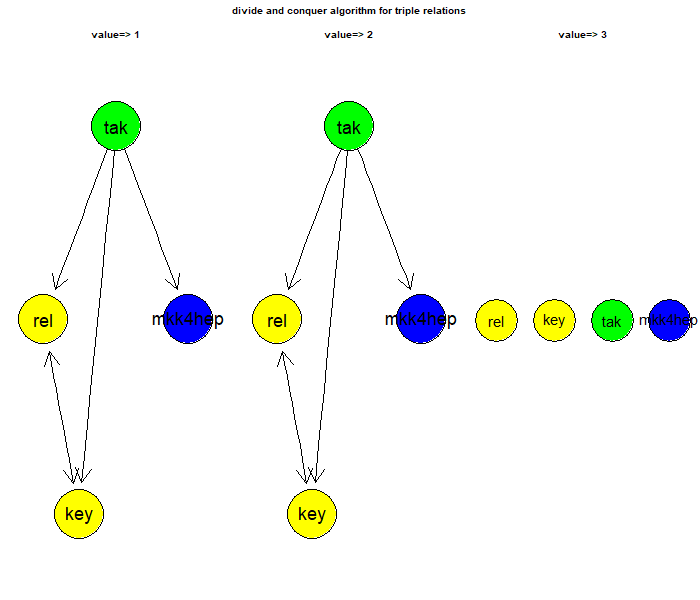
Task 2.1

The four algorithms end up with the same graph. The exhaustive search, triples, nem.greedy have the same mLL of -239.5778 the pairwise has a mLL of only -256. The selected genes are equal.

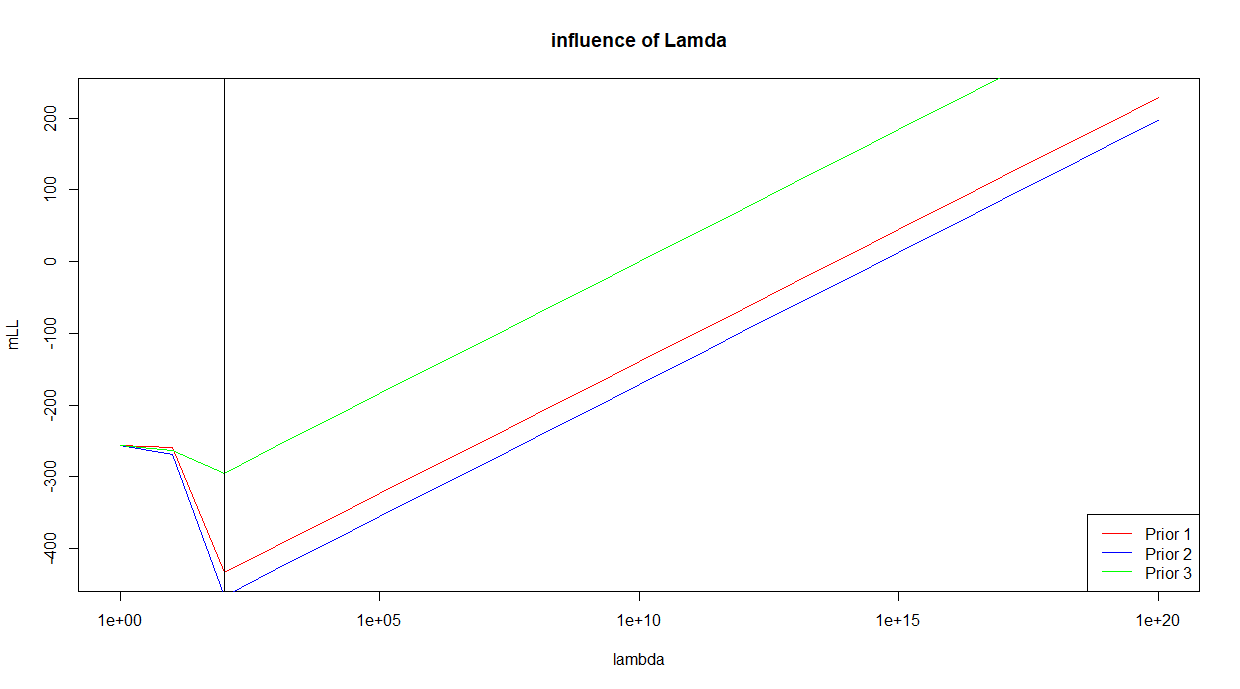


Task 2.2

For cutoff values 1,2 the model is still the same as in the tasks before. For a cutoff of at least 3 edges the graph will be fully separated.



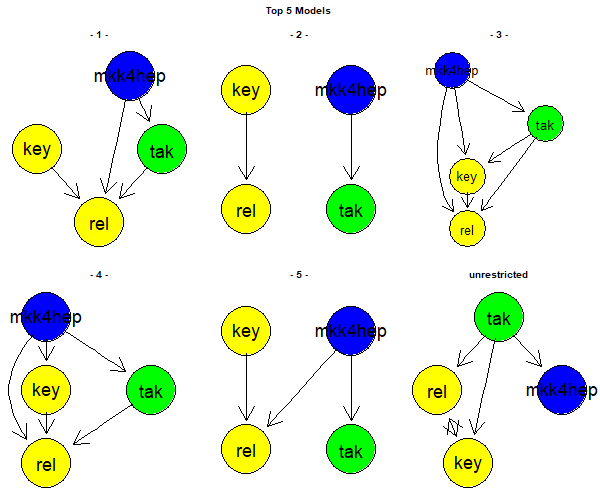
Task 3



A bigger lambda will decrease the mLL until a lamda of 100. It stepwise forces the graph to equal the prior. For lambda, bigger than 100 it seems that the mLL only depends on lambda and the data has not longer a influence to the model.

Task 4

We assume that the given directed edges contradict the edges in the opposite direction. There are 29 models which includes mkk4hep -> tak and key -> rel and excludes tak -> mkk4hep and rel -> key. We decide to just plot again the five best models.



The models look all different. But 1,3,4 are nearly equal and 2,5 similar. Model 1 is most like the unrestricted model.

Task 5

For the Boostrapfunctions we use 1000 iterations and a threshold of 0.95.

edge\_properbility

target

source rel key tak mkk4hep

rel 0.000 0.997 0.017 0.001

key 0.988 0.000 0.017 0.001

tak 0.957 0.965 0.000 0.982

mkk4hep 0.029 0.038 0.059 0.000

The models are equal to the best exhaustive search model

