BCB570 HW5

Urminder Singh

March 3, 2018

Solution 1

Bagging

Bagging is an ensembl method to improve stability of learning algorithms by reducing its variance. In general, given a dataset D with n datapoints (in any k-dimentional space), bagging generates m new datasets each of size n. It does so by doing sampling-with-replacement and samples n datapoints m times. If n is large enough, the 63.2% unique datapoints are expected to be in the m samples. A variation of bagging can have another parameter n' which is the dataset size of the sampled datasets. Here, bagging will generates m new datasets each of size n' by doing sampling-with-replacement.

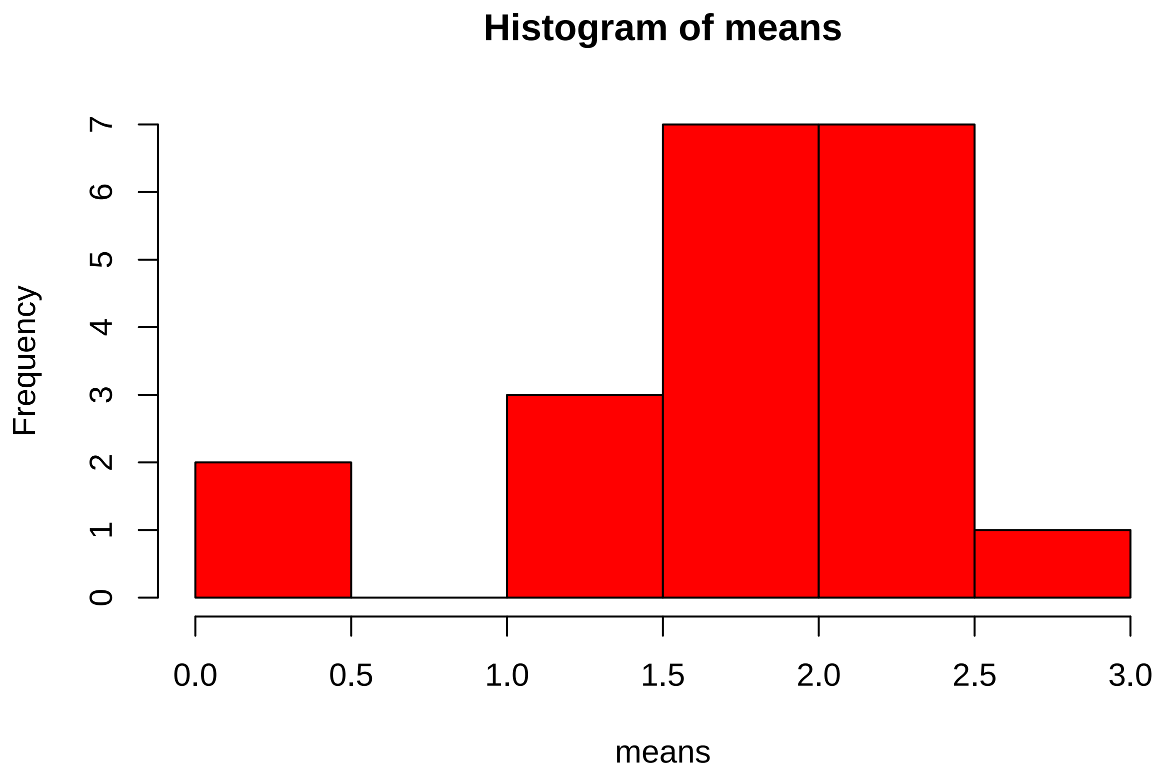
```r
# sample 100 datapoints
n_data_100 <- rnorm(100, mean = 2, sd = sqrt(8))
cat("Mean is", mean(n_data_100))
```

```
## Mean is 2.024138
```

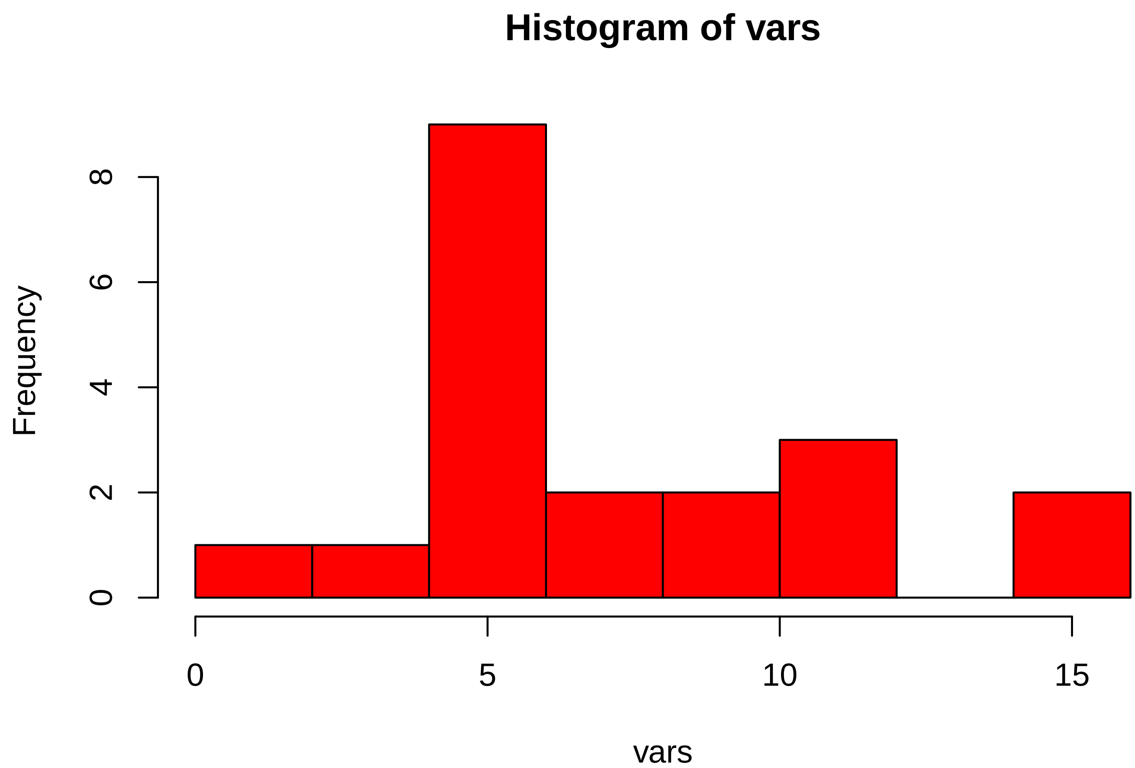
```r
cat("Var is", var(n_data_100))
```

```
## Var is 8.714753
```

```r
means <- c()
vars <- c()
N <- 20
k <- 10
for (i in 1:N) {
    this <- sample(n_data_100, k, replace = T)
    # print(this)
    means <- c(means, mean(this))
    vars <- c(vars, var(this))
}

hist(means, col = c("Red"))
```




```
hist(vars, col = c("Red"))
```




```r
e_mean <- mean(means)
e_var <- mean(vars)
```

From the histograms it is hard to estimate true parametes as we sampled very few datasets. From the sub samples the mean of means is 1.7435098 and mean of variances is 7.4471093. These are our best estimates from our sub-samples.

Solution 2

Creating association networks

To create any association network we need to go through following steps.

1. First we have to define a way that can "associate" datapoints or variables. If two points are associated we must connect them through an edge and in this way we can create a network. We have to define a mathematical function that can quantify association between two datapoints e.g. we can compute the correlation between two variables if it is greater than zero, we say the variables are accosiated.
2. We may choose an association function, depending on the type of network we want i.e directed, undirected, weighted or unweighted.E.g. if we want directed association graph our association function should not be always symmetrical.
3. After defining a function that can quantify association, we need to set a threshold such that we create edges between only the strongly associated nodes. E.g. if we define distance between two points to be our measure of association then we need to set a cut-off such that if the distance is more than a value, we do not count that as true association.

- 3a Depending on the data, sometimes it may be useful to scale all the association metrics computed to make all the association comparable with in the graph

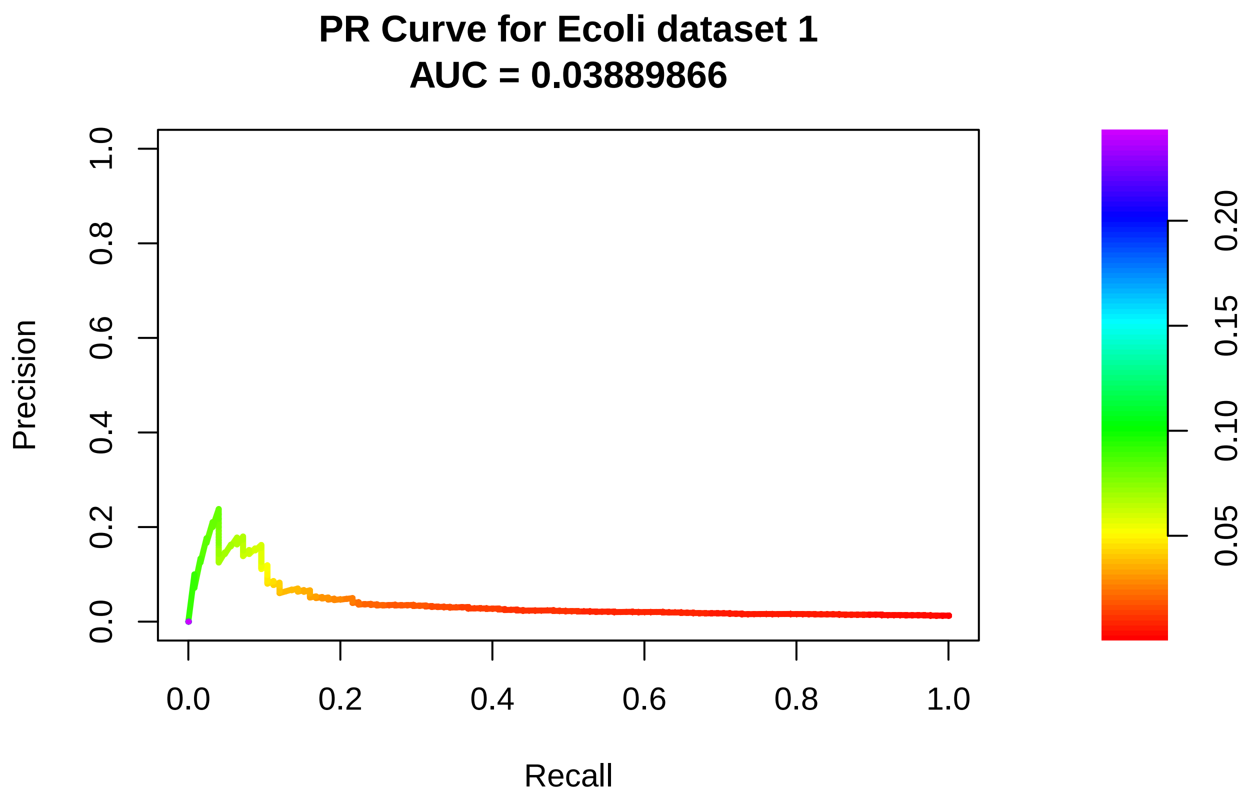
A novel association method

A method for creating association networks, not discussed in class, is using partial correlations. This method is not completely novel as it has been applied to study phenotypic networks (Chu et. al. 2014) but i am describing this in context of my data which may be novel. I am working with human transcript expression data and it would be interesting to create an association network using partial correlations. Partial correlation method finds the correlation between two variable while conditional all other variables. We can easily threshold this kind of network by calculating p-values as under the null, where all variables are independent, Hotelling is the null distribution. Using this we can test for significant partial correlations and filter the network. I can compare results of networks obtained from using correlation and partial correlation networks. A network based on partial correlation is expected to be more dense and may reveal interesting hidden links. So I will first select the transcript which are highly likely to interact using mutual information metric and filter some of the weekly interacting components. Then, on the remaining transcripts i'll use partial correlation measure to build an association network.

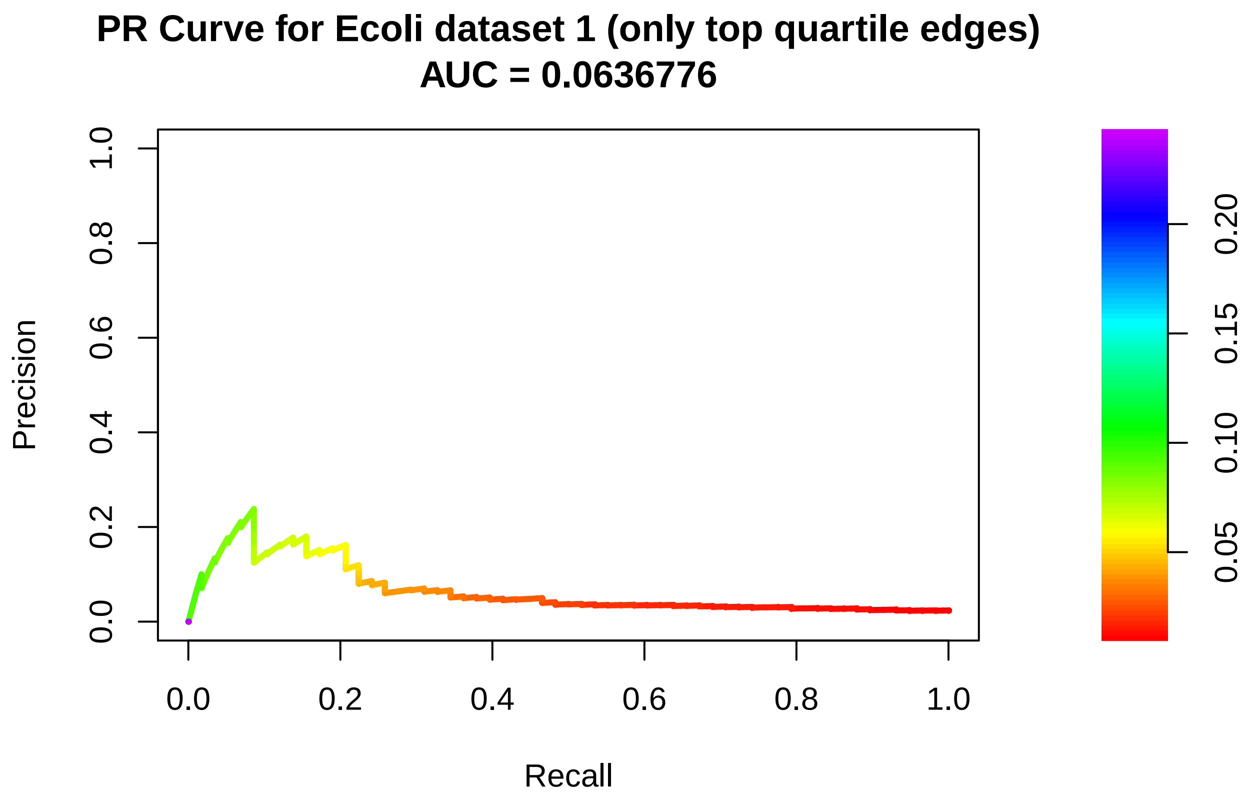
Solution 4

```
library(GENIE3)
library(readr)
library("doParallel", lib.loc = "~/R/win-library/3.4")
library("doRNG", lib.loc = "~/R/win-library/3.4")
library(plyr)
library(PRROC)
# read true labels
trueEcoli1 <- read.csv("NIHW in silico data/NIHW in silico data/Size100/DREAM3 gold standards/DREAM3Gol
    sep = "\t", header = F)
trueEcoli2 <- read.csv("NIHW in silico data/NIHW in silico data/Size100/DREAM3 gold standards/DREAM3Gol
    sep = "\t", header = F)
# do for ecoli data 1
adata <- read.csv("NIHW in silico data/NIHW in silico data/Size100/DREAM3 data/InSilicoSize100-Ecoli1-t
    sep = "\t")
adata <- t(adata)
expmat <- as.matrix(adata[2:101, 1:966])
rownames(expmat) <- paste("G", 1:100, sep = "")
colnames(expmat) <- paste("Samp", 1:966, sep = "")

res1 <- GENIE3(expmat, nCores = 6, nTrees = 100)
linkList <- getLinkList(res1)
names(linkList) <- c("RG", "TG", "Weight")
names(trueEcoli1) <- c("RG", "TG", "Label")

q4thresh <- quantile(linkList$Weight, 0.75)
# q4thresh<-0.1
linkList_q4 <- linkList[linkList$Weight >= q4thresh, ]
names(linkList_q4) <- c("RG", "TG", "Weight")
joined_data <- plyr::join(linkList, trueEcoli1)
joined_data_q4 <- plyr::join(linkList_q4, trueEcoli1)
pr1 <- pr.curve(scores.class0 = joined_data$Weight, weights.class0 = joined_data$Label,
```

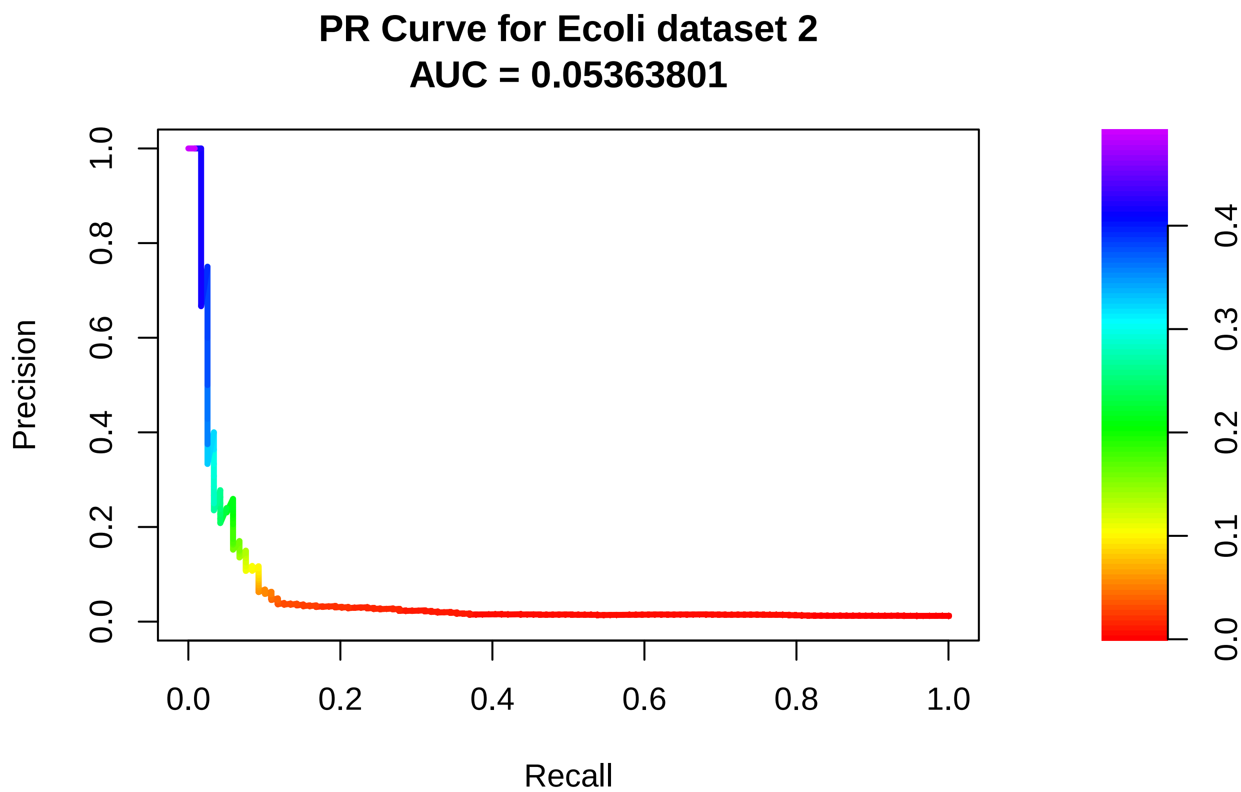
```
    curve = T)
plot(pr1, main = "PR Curve for Ecoli dataset 1")
```



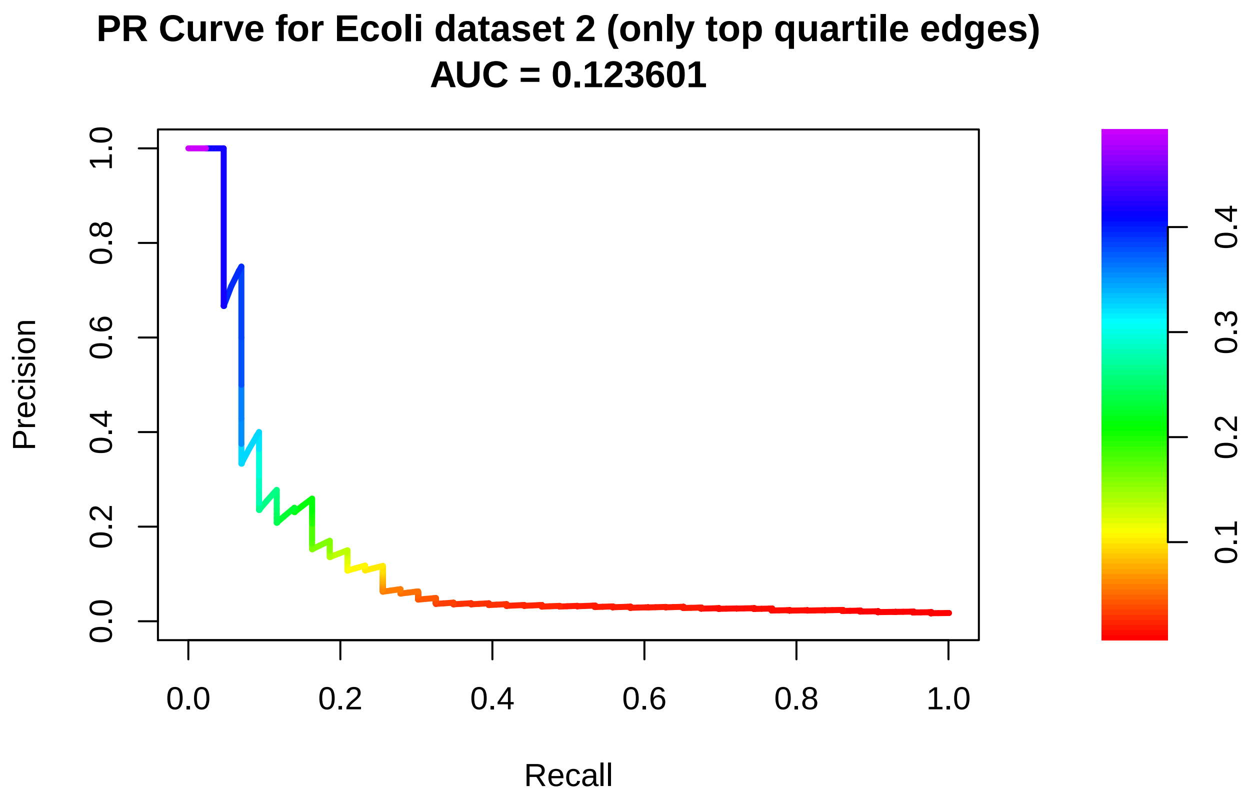

```
pr1_1 <- pr.curve(scores.class0 = joined_data_q4$Weight, weights.class0 = joined_data_q4$Label,
    curve = T)
plot(pr1_1, main = "PR Curve for Ecoli dataset 1 (only top quartile edges)")
```




```r
# do for ecoli data 2
adata2 <- read.csv("NIHW in silico data/NIHW in silico data/Size100/DREAM3 data/InSilicoSize100-Ecoli2-
    sep = "\t")
adata2 <- t(adata2)
expmat2 <- as.matrix(adata2[2:101, 1:966])
rownames(expmat2) <- paste("G", 1:100, sep = "")
colnames(expmat2) <- paste("Samp", 1:966, sep = "")

res2 <- GENIE3(expmat2, nCores = 6, K = "all", nTrees = 100)
linkList2 <- getLinkList(res2)
names(linkList2) <- c("RG", "TG", "Weight")

names(trueEcoli2) <- c("RG", "TG", "Label")
q4thresh2 <- quantile(linkList2$Weight, 0.75)
linkList2_q4 <- linkList2[linkList2$Weight >= q4thresh2, ]
names(linkList2_q4) <- c("RG", "TG", "Weight")
joined_data2 <- plyr::join(linkList2, trueEcoli2)
joined_data2_q4 <- plyr::join(linkList2_q4, trueEcoli2)
pr2 <- pr.curve(scores.class0 = joined_data2$Weight, weights.class0 = joined_data2$Label,
    curve = T)
plot(pr2, main = "PR Curve for Ecoli dataset 2")
```




```r
pr2_1 <- pr.curve(scores.class0 = joined_data2_q4$Weight, weights.class0 = joined_data2_q4$Label,
    curve = T)
plot(pr2_1, main = "PR Curve for Ecoli dataset 2 (only top quartile edges)")
```




```r
# save results to file
write_tsv(linkList, "Ecoli1_results_all.tsv")
write_tsv(linkList_q4, "Ecoli1_results_topq.tsv")
write_tsv(linkList2, "Ecoli2_results_all.tsv")
write_tsv(linkList2_q4, "Ecoli2_results_topq.tsv")

# find best estimate intersection from linkList_q4 and linkList2_q4
temp_linkList_q4 <- linkList_q4
names(temp_linkList_q4) <- c("RG", "TG", "ecoli1")
temp_linkList2_q4 <- linkList2_q4
names(temp_linkList2_q4) <- c("RG", "TG", "ecoli2")
# common edges
common <- plyr::join(temp_linkList_q4, temp_linkList2_q4, type = "inner")
# save to file
write_tsv(common, "Ecoli_best_estimate.tsv")
```

a

GENIE3 is an algorithm for inferring gene regulatory networks from expression data. GENIE3 decomposes the prediction of a regulatory network between p genes into p different regression problems. In each of the regression problems, the expression pattern of one of the genes (target gene) is predicted from the expression patterns of all the other genes (input genes), using tree-based ensemble methods.

I ran GENEI3 with default parameters except I changed number of trees to 100 (to be faster). To threshold the network I took the top quartile values and selected the edges.

b

I ran GENEI3 for the two ecoli data sets on the trajectories file. Then I plotted the PR curve as shown in the attached figures. Then I filtered my networks by keeping only the edges which have weights in the top quartile. To come up with my best estimate about Ecoli network, I took the intersection of the two results i.e. if edge is present in the two networks. I saved these results to file “Ecoli_best_estimate.tsv”.

c Best estimates for ecoli data sets

```
common
```

```
##       RG  TG     ecoli1      ecoli2
## 1    G66 G68 0.09550842 0.033417604
## 2    G18 G16 0.08563683 0.011648596
## 3    G31 G30 0.08346623 0.040799943
## 4    G68 G66 0.08264686 0.016296007
## 5    G37 G93 0.07636189 0.064046040
## 6    G28 G45 0.07032582 0.015785393
## 7    G63 G60 0.06979220 0.022103001
## 8    G18 G32 0.06965476 0.016845274
## 9    G16 G18 0.06771008 0.010163507
## 10   G14 G93 0.06759043 0.020017658
## 11   G41 G45 0.06751179 0.009864705
## 12   G18 G46 0.06656073 0.024081322
## 13   G99  G5 0.06433153 0.010759347
## 14   G82 G43 0.06365858 0.019729160
## 15   G45 G28 0.06185529 0.019163994
## 16   G30 G31 0.06135637 0.030019798
## 17   G40  G5 0.06094871 0.016461434
## 18   G93 G37 0.06074766 0.098612802
## 19   G31 G36 0.06066831 0.011496761
## 20   G97 G39 0.05814637 0.009990772
## 21   G83 G38 0.05740563 0.036520452
## 22   G32 G18 0.05688047 0.012798603
## 23   G90 G60 0.05641278 0.010744575
## 24   G88 G36 0.05596317 0.009386102
## 25   G99  G2 0.05566076 0.041797108
## 26   G33 G20 0.05548770 0.033741731
## 27   G23 G81 0.05509880 0.016668972
## 28   G40 G17 0.05333692 0.030176309
## 29   G36 G34 0.05256332 0.009601923
## 30   G23 G21 0.05243834 0.011945607
## 31   G16  G9 0.05236794 0.040137635
## 32    G5 G49 0.05168572 0.022957553
## 33   G17  G5 0.05078428 0.009215007
## 34   G18 G59 0.04987122 0.009471102
## 35   G51 G87 0.04965906 0.011863316
## 36   G63 G64 0.04945444 0.171017427
## 37   G16 G80 0.04938229 0.019806393
## 38   G28 G98 0.04801344 0.010306426
## 39   G82 G83 0.04671165 0.417102096
## 40   G47 G50 0.04428631 0.009279675
## 41   G51  G8 0.04387378 0.044331969
```

```
## 42   G61  G59 0.04380656 0.009579306
## 43   G61  G64 0.04361619 0.019268455
## 44   G39  G38 0.04357242 0.015530803
## 45   G32  G59 0.04262978 0.018944100
## 46   G68  G55 0.04177970 0.011411278
## 47   G50  G47 0.04176139 0.010402571
## 48   G43  G97 0.04149473 0.014574427
## 49    G5  G85 0.04105565 0.009799214
## 50   G64  G63 0.04093342 0.105197477
## 51   G14  G45 0.04081945 0.011576487
## 52   G75   G3 0.04020607 0.042214776
## 53   G51  G40 0.04012734 0.019246157
## 54   G68  G70 0.04011311 0.013675525
## 55   G49   G5 0.03998728 0.012354642
## 56   G41   G2 0.03975259 0.016928342
## 57   G39  G93 0.03972749 0.011024929
## 58   G97   G2 0.03968363 0.041735961
## 59   G16  G20 0.03955842 0.011896984
## 60   G17   G8 0.03920944 0.010039812
## 61   G41  G16 0.03914060 0.010129431
## 62   G35  G26 0.03912696 0.016311448
## 63   G80   G9 0.03889456 0.011233662
## 64   G21  G79 0.03876957 0.011618636
## 65    G9  G23 0.03849022 0.010621985
## 66   G49  G31 0.03808134 0.012153970
## 67   G64  G41 0.03797340 0.023641587
## 68   G34  G36 0.03792750 0.010155900
## 69   G60  G57 0.03790503 0.095877730
## 70   G40  G80 0.03786658 0.034194037
## 71   G82  G12 0.03776867 0.009234299
## 72   G99  G63 0.03732218 0.010507367
## 73   G21  G23 0.03703495 0.010749170
## 74   G41  G47 0.03692703 0.030236280
## 75   G16  G32 0.03689561 0.016269992
## 76   G49  G96 0.03677504 0.009815624
## 77   G23   G9 0.03674412 0.031960401
## 78   G28  G30 0.03673785 0.017631801
## 79   G37  G18 0.03618648 0.013782284
## 80   G81  G86 0.03616645 0.018829778
## 81   G23  G59 0.03605691 0.010950341
## 82   G93  G41 0.03603287 0.061618257
## 83   G17   G6 0.03581804 0.020602758
## 84   G93  G40 0.03560491 0.117911092
## 85   G80  G40 0.03547144 0.011042922
## 86   G64  G59 0.03542134 0.009930045
## 87   G96  G49 0.03534876 0.030223350
## 88   G14  G86 0.03530097 0.011460183
## 89   G18  G37 0.03527731 0.010415026
## 90   G34  G19 0.03491836 0.020757611
## 91   G55  G44 0.03477841 0.010134641
## 92    G3  G75 0.03469260 0.024129155
## 93   G88  G86 0.03431062 0.014521874
## 94   G17  G40 0.03421996 0.077210727
## 95   G46  G18 0.03365283 0.016662045
```

```
## 96   G32  G38 0.03359961 0.023757966
## 97   G42  G30 0.03359512 0.026147607
## 98    G2  G95 0.03353169 0.022873042
## 99   G47  G92 0.03340733 0.014101340
## 100  G64  G61 0.03318244 0.024101959
## 101  G92  G44 0.03308052 0.013338764
## 102   G6  G94 0.03297438 0.019151756
## 103  G72  G73 0.03273757 0.023934662
## 104  G81  G23 0.03215453 0.036361903
## 105  G37  G61 0.03204300 0.053511280
## 106  G81   G2 0.03199372 0.016084251
## 107  G55  G89 0.03194773 0.014502308
## 108  G79   G5 0.03165494 0.010691902
## 109  G42  G99 0.03151784 0.060846343
## 110  G40   G6 0.03151063 0.084998087
## 111  G17  G42 0.03147548 0.280873963
## 112  G52  G76 0.03142608 0.013721138
## 113  G13  G42 0.03112532 0.011471305
## 114  G18  G88 0.03067207 0.016998158
## 115  G46  G79 0.03061646 0.009568428
## 116  G79  G88 0.03058667 0.014618820
## 117  G16  G79 0.03045550 0.037622101
## 118  G37   G2 0.03045396 0.009204104
## 119  G25  G26 0.03039506 0.016177434
## 120   G5  G79 0.03018807 0.010922002
## 121   G8  G46 0.03013462 0.016400182
## 122  G39  G97 0.02995462 0.020425977
## 123  G37  G41 0.02990516 0.011930989
## 124  G51  G13 0.02988311 0.017857681
## 125  G11  G29 0.02977565 0.009491852
## 126  G80  G60 0.02961760 0.019747560
## 127  G18  G97 0.02949718 0.032383359
## 128  G58  G13 0.02949252 0.009288736
## 129  G43  G24 0.02948547 0.018347530
## 130  G96  G44 0.02946444 0.010797919
## 131  G83  G82 0.02941753 0.418905835
## 132  G40  G41 0.02913226 0.010459206
## 133  G37  G90 0.02890624 0.011492933
## 134  G42  G20 0.02849433 0.018953389
## 135  G80  G49 0.02846486 0.016948099
## 136  G41  G93 0.02839658 0.038047106
## 137  G24   G5 0.02818025 0.019503890
## 138  G18  G80 0.02813341 0.011304061
## 139  G37  G48 0.02812036 0.017863032
## 140  G16  G23 0.02796637 0.009208941
## 141  G69  G78 0.02790849 0.015470520
## 142  G86   G8 0.02783405 0.010243700
## 143  G58  G45 0.02773012 0.014003509
## 144  G32   G4 0.02772748 0.012981682
## 145  G39   G2 0.02757343 0.009158128
## 146  G76  G46 0.02743362 0.057231123
## 147  G99  G42 0.02743241 0.010492830
## 148  G60  G90 0.02740925 0.014537633
## 149  G18  G41 0.02740297 0.012614910
```

```
## 150  G40  G51 0.02727947 0.012082339
## 151  G86  G90 0.02725303 0.011844617
## 152  G62  G57 0.02722220 0.019749474
## 153  G79  G16 0.02698078 0.031840512
## 154  G15   G3 0.02684842 0.012039795
## 155  G40  G24 0.02673968 0.043076456
## 156  G86  G46 0.02666464 0.013304635
## 157   G9  G27 0.02665807 0.023369549
## 158  G23  G20 0.02658056 0.020575826
## 159  G67  G22 0.02654312 0.014760255
## 160  G37  G96 0.02650491 0.062020710
## 161  G55  G68 0.02626418 0.023233419
## 162  G97   G4 0.02620511 0.022084775
## 163  G38  G10 0.02618078 0.104777401
## 164  G49  G94 0.02610108 0.056201351
## 165   G6  G46 0.02598914 0.009401715
## 166  G99  G11 0.02595309 0.157582516
## 167  G38   G1 0.02584902 0.015432025
## 168   G8  G51 0.02582904 0.010972166
## 169   G6  G40 0.02578767 0.082968102
## 170  G57  G44 0.02565458 0.009182807
## 171  G23  G30 0.02564940 0.012160931
## 172  G81  G14 0.02553622 0.028191590
## 173  G88  G16 0.02545079 0.024796099
## 174  G99  G61 0.02542064 0.253702689
## 175  G87  G85 0.02541927 0.023492212
## 176  G88  G53 0.02538553 0.009763460
## 177  G57  G68 0.02526282 0.010101693
## 178  G32  G16 0.02524275 0.019783422
## 179  G46  G89 0.02521256 0.056506941
## 180  G95  G76 0.02503465 0.016012124
## 181  G77  G44 0.02499024 0.017505185
## 182  G14  G97 0.02492174 0.018670561
## 183  G24   G7 0.02487888 0.024935121
## 184 G100  G35 0.02467739 0.018547887
## 185  G39  G53 0.02455833 0.027538762
## 186  G47  G83 0.02442998 0.016769833
## 187  G62  G15 0.02439291 0.009463404
## 188  G81  G85 0.02435790 0.034181180
## 189  G61  G37 0.02434209 0.119170130
## 190  G20  G84 0.02432673 0.047375161
## 191  G83  G24 0.02425830 0.017046953
## 192  G40  G90 0.02416804 0.020720708
## 193  G59  G32 0.02413861 0.014546624
## 194  G42  G18 0.02411510 0.010119378
## 195  G42  G17 0.02401998 0.026920027
## 196  G69  G55 0.02401403 0.010038204
## 197  G43  G30 0.02382991 0.019871617
## 198  G76   G8 0.02374562 0.010659387
## 199   G4  G21 0.02366010 0.021214416
## 200  G88  G99 0.02340800 0.009097908
## 201  G94  G27 0.02328728 0.009612790
## 202  G61  G96 0.02315969 0.011370099
## 203  G90  G37 0.02297018 0.010455004
```

```
## 204   G19   G1 0.02292810 0.009300770
## 205   G86  G81 0.02291787 0.027262685
## 206   G22  G98 0.02273926 0.019581942
## 207   G32  G97 0.02266802 0.012611230
## 208    G5   G2 0.02263099 0.022866569
## 209   G32  G14 0.02251653 0.009154635
## 210   G12  G82 0.02241730 0.018755391
## 211   G79   G9 0.02241616 0.010028960
## 212 G100  G26 0.02236407 0.010134460
## 213   G19  G34 0.02222290 0.011776330
## 214   G20  G33 0.02207739 0.043474912
## 215   G98  G22 0.02195363 0.012131420
## 216    G4  G97 0.02182152 0.015194895
## 217   G13  G58 0.02165284 0.013805584
## 218   G72  G30 0.02162316 0.019717225
## 219   G17   G9 0.02154795 0.012761838
## 220   G87  G26 0.02144717 0.011536872
## 221   G65  G99 0.02142142 0.009535085
## 222   G16  G21 0.02138926 0.016677566
## 223   G77  G90 0.02130657 0.011046409
## 224   G14   G4 0.02124634 0.010807627
## 225   G22  G33 0.02115101 0.021117472
## 226   G39  G82 0.02105678 0.012661378
## 227   G64  G88 0.02105601 0.024356623
## 228   G57  G61 0.02104253 0.067656961
## 229   G30  G45 0.02098128 0.009771188
## 230   G16  G82 0.02096476 0.012933254
## 231   G49  G86 0.02095358 0.009139450
## 232   G39  G30 0.02093603 0.009215716
## 233    G9  G16 0.02081892 0.040178600
## 234   G16  G88 0.02066553 0.011588401
## 235   G65  G68 0.02065377 0.049096120
## 236   G86  G30 0.02063648 0.069965405
## 237   G16  G38 0.02060223 0.028007603
## 238    G6  G17 0.02055446 0.021048772
## 239   G16  G36 0.02047100 0.010885681
## 240   G24  G83 0.02044581 0.016858198
## 241   G49  G85 0.02038770 0.012256362
## 242   G93  G90 0.02038068 0.017469963
## 243   G40  G47 0.02034198 0.141261045
## 244   G47  G56 0.02026722 0.177109075
## 245   G85  G35 0.02018549 0.011140317
## 246   G40  G93 0.02010592 0.177495622
## 247   G38  G97 0.02009619 0.036925234
## 248   G31  G92 0.02008554 0.012237884
## 249   G46  G82 0.02008386 0.015844668
## 250   G15  G62 0.02001203 0.009740097
## 251   G84  G78 0.02000378 0.012310049
## 252    G5  G65 0.01979031 0.009137950
## 253   G65   G3 0.01974739 0.027304906
## 254   G75  G10 0.01970993 0.013818826
## 255   G96  G37 0.01961407 0.022752283
## 256   G28  G58 0.01945867 0.053877396
## 257   G45  G19 0.01943275 0.013252903
```

```
## 258  G88  G21 0.01940122 0.018776731
## 259  G35  G25 0.01939413 0.014079186
## 260  G67  G34 0.01936812 0.010128516
## 261  G28  G95 0.01936608 0.012192983
## 262  G55  G22 0.01923017 0.009782407
## 263  G37  G92 0.01912600 0.044517556
## 264  G67  G10 0.01910882 0.010617525
## 265  G51   G6 0.01906539 0.010624452
## 266  G68  G33 0.01883160 0.009982445
## 267  G99  G45 0.01880047 0.017958120
## 268  G44  G54 0.01871646 0.013526382
## 269  G61  G75 0.01863092 0.012499654
## 270  G71  G91 0.01857641 0.041809929
## 271  G68  G94 0.01850769 0.013428033
## 272  G45  G30 0.01840904 0.010668446
## 273  G38  G36 0.01840693 0.023432195
## 274  G99 G100 0.01840239 0.048980318
## 275  G22   G3 0.01837965 0.010503218
## 276  G60  G62 0.01836910 0.490542659
## 277  G83  G40 0.01834571 0.011823535
## 278  G47   G5 0.01828994 0.019559381
## 279  G51  G57 0.01822507 0.099632801
## 280  G39  G24 0.01814695 0.023800626
## 281  G97  G38 0.01813050 0.015277927
## 282  G42  G64 0.01806440 0.027929008
## 283  G32  G43 0.01803007 0.009973288
## 284  G88  G33 0.01798038 0.010835917
## 285  G17  G29 0.01786064 0.097107139
## 286  G86  G58 0.01785409 0.025114301
## 287  G31  G49 0.01783950 0.019322154
## 288  G16  G47 0.01770143 0.019120368
## 289  G31  G85 0.01768730 0.025037921
## 290  G24  G40 0.01768725 0.031242118
## 291  G96  G36 0.01767627 0.009622829
## 292  G42  G70 0.01760042 0.011465829
## 293  G28  G44 0.01753973 0.047949126
## 294  G90  G48 0.01751897 0.016806665
## 295  G59  G95 0.01748908 0.013600371
## 296  G67  G11 0.01741636 0.017026702
## 297  G26  G32 0.01740608 0.016375619
## 298  G83  G11 0.01740359 0.014736847
## 299  G94  G22 0.01737641 0.009509881
## 300  G69  G98 0.01733688 0.016012438
## 301  G43  G19 0.01727403 0.018306991
## 302  G38  G45 0.01727314 0.009411491
## 303  G46   G5 0.01726559 0.012882083
## 304  G27  G94 0.01725614 0.011002777
## 305  G83  G92 0.01723145 0.020738453
## 306  G71  G94 0.01721411 0.047587003
## 307  G18  G61 0.01718906 0.011624822
## 308   G5   G4 0.01717228 0.009567965
## 309  G94  G32 0.01716044 0.017950325
## 310  G18  G36 0.01709799 0.036432856
## 311  G56  G45 0.01708312 0.010860853
```

```
## 312  G35   G11 0.01704876 0.011520940
## 313  G48   G10 0.01703618 0.010406531
## 314  G29  G100 0.01703143 0.010128672
## 315  G82   G38 0.01701001 0.012442141
## 316  G48   G90 0.01700668 0.009955744
## 317  G53   G39 0.01693261 0.014702219
## 318  G41    G8 0.01678702 0.022236942
## 319  G41   G18 0.01675363 0.013116313
## 320   G3   G15 0.01671301 0.010452184
## 321   G5   G20 0.01663503 0.011419601
## 322  G42    G9 0.01660603 0.010819614
## 323   G5   G97 0.01656814 0.020171757
## 324  G31   G35 0.01649774 0.019609281
## 325  G61   G66 0.01649005 0.018533396
## 326  G38   G16 0.01647612 0.013182628
## 327  G68   G18 0.01641125 0.010855976
## 328  G88   G79 0.01638413 0.020097724
## 329  G93   G78 0.01636164 0.011032707
## 330  G85   G31 0.01635956 0.017563262
## 331  G29    G7 0.01634800 0.012765010
## 332  G24   G43 0.01631628 0.024979444
## 333  G23   G85 0.01627341 0.013681186
## 334  G21   G12 0.01626646 0.023751985
## 335 G100   G19 0.01626360 0.010517150
## 336  G96   G79 0.01625237 0.009406822
## 337  G42   G79 0.01623063 0.013327288
## 338  G56    G3 0.01617687 0.012521781
## 339   G8   G43 0.01617093 0.011894870
## 340  G11   G99 0.01617035 0.015549491
## 341  G75    G4 0.01616822 0.012450542
## 342  G91   G71 0.01616065 0.016870285
## 343  G16    G5 0.01614715 0.009189204
## 344  G10    G4 0.01613059 0.015743323
## 345  G94   G23 0.01613015 0.017069224
## 346  G30   G72 0.01611574 0.053427944
## 347  G53    G8 0.01607181 0.011073219
## 348   G6   G99 0.01601490 0.011450286
## 349  G28   G55 0.01598590 0.181649580
## 350  G50   G73 0.01598370 0.010864380
## 351  G40   G83 0.01597973 0.013505534
## 352  G99   G14 0.01595113 0.049069144
## 353  G86   G48 0.01589628 0.028648161
## 354  G20   G42 0.01580886 0.035178734
## 355  G68   G65 0.01577425 0.041717799
## 356  G15    G2 0.01576859 0.017831599
## 357   G9   G41 0.01574439 0.010188184
## 358  G85    G7 0.01570192 0.023189832
## 359  G40    G2 0.01569180 0.013858437
## 360  G41   G69 0.01565974 0.013018431
## 361  G77   G37 0.01562581 0.012535747
## 362  G51   G52 0.01561638 0.010140027
## 363   G5   G37 0.01560347 0.014625046
## 364  G46   G31 0.01559687 0.029798328
## 365  G23   G94 0.01556769 0.020246974
```

```
## 366   G84    G9 0.01553838 0.010527548
## 367   G13   G51 0.01551827 0.011081165
## 368   G85    G5 0.01549825 0.015431244
## 369   G12   G35 0.01545983 0.013404229
## 370   G37   G47 0.01545250 0.045107315
## 371    G7   G55 0.01543208 0.010240296
## 372   G45    G1 0.01542566 0.009488704
## 373   G86   G49 0.01530564 0.033525334
## 374   G87  G100 0.01524991 0.030413984
## 375    G1   G26 0.01523001 0.009098830
## 376   G93   G88 0.01522897 0.015783968
## 377   G38    G7 0.01518032 0.033588212
## 378    G2   G97 0.01510503 0.023785365
## 379   G37   G33 0.01507438 0.018372747
## 380   G79   G96 0.01503952 0.011882336
## 381   G22   G95 0.01501729 0.009441269
## 382   G41   G86 0.01500967 0.019368219
## 383   G39   G18 0.01496589 0.011467202
## 384   G93   G54 0.01495959 0.040548213
## 385   G81   G33 0.01494285 0.010115001
## 386   G11   G94 0.01485882 0.014269840
## 387   G90   G21 0.01484396 0.017527438
## 388   G75   G74 0.01483815 0.014888281
## 389   G21   G44 0.01483484 0.011354653
## 390   G61   G99 0.01481441 0.111345677
## 391   G47   G45 0.01475194 0.013071939
## 392   G42    G6 0.01474275 0.016596194
## 393   G98    G9 0.01473430 0.052290652
## 394   G42   G49 0.01473216 0.023939878
## 395   G12   G29 0.01472886 0.055399138
## 396   G52   G98 0.01470417 0.012008822
## 397   G19   G77 0.01470156 0.013076185
## 398   G25   G40 0.01469556 0.118798319
## 399   G26   G75 0.01466622 0.009254634
## 400   G83    G6 0.01465083 0.022116737
## 401   G80   G96 0.01460888 0.056622411
## 402   G49   G87 0.01460478 0.010889459
## 403   G96   G85 0.01459812 0.114418078
## 404   G28   G35 0.01453037 0.029399528
## 405   G18   G73 0.01452774 0.009093985
## 406   G95   G54 0.01449030 0.009172992
## 407   G46   G22 0.01446725 0.010699880
## 408   G81   G93 0.01446226 0.011179217
## 409   G17   G98 0.01444894 0.021129048
## 410   G43   G84 0.01444864 0.061568161
## 411  G100   G87 0.01442141 0.028906691
## 412   G81   G36 0.01442040 0.016381010
## 413   G10   G38 0.01438529 0.113558351
## 414   G61   G57 0.01434348 0.133824163
## 415   G19   G12 0.01427479 0.009295462
## 416   G55   G28 0.01426799 0.159091007
## 417   G70   G95 0.01425885 0.016175884
## 418   G11   G67 0.01419555 0.016403284
## 419   G90    G7 0.01417693 0.010390860
```

```
## 420  G41  G95 0.01414471 0.015615795
## 421  G21  G55 0.01414185 0.014689030
## 422  G32  G94 0.01413783 0.025243073
## 423  G57  G53 0.01407677 0.018279446
## 424  G82  G35 0.01407298 0.011527240
## 425   G2  G65 0.01406751 0.009445141
## 426  G50  G32 0.01405308 0.011462763
## 427  G72  G95 0.01403188 0.010007607
## 428   G8  G41 0.01402477 0.014535423
## 429  G99   G6 0.01402297 0.024165197
## 430  G16  G49 0.01399440 0.011428542
## 431  G26   G3 0.01394183 0.018866989
## 432  G94  G38 0.01393746 0.009729380
## 433  G45  G14 0.01386169 0.009625025
## 434  G84  G43 0.01377998 0.125070162
## 435  G43  G90 0.01377740 0.010233739
## 436  G95  G22 0.01375573 0.012812126
## 437  G36  G73 0.01374495 0.011500078
## 438  G32  G26 0.01368723 0.023958688
## 439  G40  G29 0.01367715 0.014896975
## 440  G43  G55 0.01367136 0.010550320
## 441  G60  G19 0.01366761 0.009368599
## 442  G59  G68 0.01365038 0.012601427
## 443  G88  G18 0.01364628 0.016671784
## 444  G87  G57 0.01361795 0.015294108
## 445  G81  G57 0.01345910 0.011521366
## 446  G11  G75 0.01343301 0.034559411
## 447  G80  G30 0.01338782 0.010222176
## 448  G60  G30 0.01338250 0.012412447
## 449  G59  G15 0.01338158 0.010740322
## 450  G32  G20 0.01338003 0.012928830
## 451  G61   G7 0.01334976 0.012792435
## 452  G58  G75 0.01333593 0.009302470
## 453  G14  G99 0.01328851 0.021292707
## 454  G21  G16 0.01322350 0.011323120
## 455  G47  G41 0.01321562 0.036365825
## 456  G86  G88 0.01321036 0.027742974
## 457  G25  G35 0.01317762 0.014967649
## 458  G13  G77 0.01316635 0.038919606
## 459  G97   G9 0.01315402 0.009210860
## 460  G87  G33 0.01313611 0.018300044
## 461  G17  G25 0.01311419 0.325484980
## 462  G26  G74 0.01308281 0.009638134
## 463  G64   G6 0.01304450 0.021379716
## 464  G44  G86 0.01303411 0.010007935
## 465  G87  G88 0.01303317 0.016424158
## 466  G21   G4 0.01299501 0.010667449
## 467  G18  G68 0.01294368 0.014234630
## 468  G34  G65 0.01293528 0.024850227
## 469  G50  G60 0.01293454 0.072773834
## 470 G100  G29 0.01292581 0.020570792
## 471  G64  G90 0.01291201 0.011138635
## 472  G97  G32 0.01290020 0.009282077
## 473  G41  G99 0.01289375 0.020234334
```

```
## 474  G54  G69 0.01285808 0.011157360
## 475  G78  G75 0.01282698 0.024284713
## 476  G96   G5 0.01281734 0.053446627
## 477  G86   G7 0.01281718 0.014462333
## 478  G76  G95 0.01280227 0.024591993
## 479  G35 G100 0.01280102 0.018013066
## 480  G25  G86 0.01278824 0.018993770
## 481   G3  G98 0.01277875 0.016666734
## 482   G5  G96 0.01276398 0.010420387
## 483  G73  G22 0.01273447 0.049359132
## 484  G24   G8 0.01268931 0.011001013
## 485  G49  G54 0.01268750 0.011084513
## 486  G35  G76 0.01267817 0.010945475
## 487  G42  G56 0.01266410 0.122406268
## 488  G23  G98 0.01265348 0.009966886
## 489  G97  G60 0.01264488 0.015374102
## 490   G5  G19 0.01262581 0.010131271
## 491  G27   G9 0.01262388 0.021090441
## 492  G53  G72 0.01260598 0.010837525
## 493  G55   G7 0.01257086 0.011098057
## 494  G22  G77 0.01256433 0.009695280
## 495  G24  G55 0.01255827 0.018003745
## 496  G65  G26 0.01255357 0.018274407
## 497   G6  G59 0.01255183 0.020819578
## 498  G71  G97 0.01250104 0.010743000
## 499  G63   G4 0.01246245 0.011394948
## 500  G85   G1 0.01243102 0.013259278
## 501  G87  G22 0.01242173 0.021061409
## 502  G94  G11 0.01238433 0.012199998
## 503  G46  G19 0.01238305 0.016645259
## 504  G13  G78 0.01238154 0.021196908
## 505  G22  G35 0.01235661 0.010876488
## 506  G15  G65 0.01234151 0.043810183
## 507   G9  G49 0.01230266 0.016856883
## 508  G98  G77 0.01229857 0.011133065
## 509  G78  G76 0.01229598 0.017456808
## 510  G99  G18 0.01228614 0.011249321
## 511  G42  G71 0.01224624 0.024090593
## 512  G68  G19 0.01223817 0.016180475
## 513  G47  G53 0.01221548 0.027245233
## 514   G5  G36 0.01218814 0.018602665
## 515   G8  G74 0.01218394 0.155015810
## 516  G96  G80 0.01216915 0.048234874
## 517  G65  G34 0.01212646 0.017640514
## 518  G16  G26 0.01209769 0.009709399
## 519  G63  G61 0.01209747 0.025714881
## 520  G13  G52 0.01208915 0.032781430
## 521  G38  G87 0.01208392 0.010649580
## 522  G54  G32 0.01207869 0.018169013
## 523  G19  G95 0.01204624 0.014157374
## 524   G4  G32 0.01203124 0.026639951
## 525  G28  G33 0.01199002 0.011286573
## 526  G72  G55 0.01197957 0.032077539
## 527  G57  G90 0.01197922 0.013947796
```

```
## 528   G9  G97 0.01196357 0.016172742
## 529  G30  G42 0.01196186 0.014151602
## 530  G65  G67 0.01195916 0.022287479
## 531   G1  G25 0.01195531 0.018248083
## 532   G9  G80 0.01195208 0.031935122
## 533  G12  G68 0.01194153 0.010683526
## 534  G46  G78 0.01194090 0.013394220
## 535   G9  G79 0.01193852 0.014879857
## 536  G51  G44 0.01190971 0.010365370
## 537  G52  G69 0.01190916 0.046502161
## 538  G58  G28 0.01190019 0.036167119
## 539   G4  G34 0.01189736 0.017816258
## 540  G14  G79 0.01188854 0.011826651
## 541  G57  G60 0.01188579 0.178531193
## 542  G54  G52 0.01186704 0.030196421
## 543  G59  G76 0.01186342 0.011738710
## 544  G71  G74 0.01186058 0.202565284
## 545  G67  G65 0.01185456 0.055414176
## 546   G5  G30 0.01184310 0.009084817
## 547  G42  G98 0.01183430 0.009586116
## 548  G24  G51 0.01178964 0.094530260
## 549   G7  G12 0.01178885 0.013392963
## 550  G25  G29 0.01178818 0.153889994
## 551   G1  G95 0.01177757 0.011857955
## 552  G67  G72 0.01177642 0.122266908
## 553  G27  G38 0.01177181 0.009919759
## 554  G29  G25 0.01174107 0.075111003
## 555  G65  G89 0.01174094 0.030125698
## 556  G56  G27 0.01166361 0.011876149
## 557  G17  G81 0.01165990 0.054889798
## 558  G98  G33 0.01165531 0.011394872
## 559  G45  G99 0.01165213 0.015756204
## 560  G69  G33 0.01163671 0.010043605
## 561  G69  G12 0.01163168 0.017258974
## 562  G29   G9 0.01162484 0.011359451
## 563  G90  G20 0.01160649 0.009430369
## 564  G25   G2 0.01160560 0.013343176
## 565  G57  G29 0.01159388 0.013165255
## 566  G50   G6 0.01158999 0.009595716
## 567  G44  G28 0.01158147 0.031678435
## 568  G22  G13 0.01154797 0.014924112
## 569  G85  G20 0.01154639 0.017283133
## 570  G98  G48 0.01154057 0.025002217
## 571  G93  G30 0.01152947 0.009394704
## 572  G81  G97 0.01151711 0.013524157
## 573  G15  G66 0.01150310 0.015762558
## 574  G31  G70 0.01148153 0.015493325
## 575   G5  G88 0.01147999 0.016077248
## 576  G56   G7 0.01147874 0.009616466
## 577  G57  G54 0.01147032 0.009470275
## 578  G83  G58 0.01145307 0.027721954
## 579  G90  G56 0.01141404 0.010330381
## 580  G77  G19 0.01141376 0.031585800
## 581  G34  G33 0.01141181 0.023626918
```

```
## 582  G85  G87 0.01140138 0.009911849
## 583  G98  G19 0.01139896 0.027799950
## 584  G85  G49 0.01139823 0.018906174
## 585   G6  G22 0.01139788 0.012500515
## 586  G56  G74 0.01137856 0.045079057
## 587  G81  G58 0.01137696 0.024295348
## 588  G26  G67 0.01137115 0.011478655
## 589  G32  G72 0.01135161 0.012024176
## 590  G21  G76 0.01134108 0.012540520
## 591  G29  G12 0.01133755 0.069859768
## 592  G10  G34 0.01133539 0.020846839
## 593  G67  G33 0.01130666 0.070128155
## 594  G98  G75 0.01130464 0.013704869
## 595  G96  G86 0.01129394 0.015259079
## 596  G42  G73 0.01128977 0.053787052
## 597  G40  G25 0.01128472 0.022844993
## 598  G55   G4 0.01127368 0.012096818
## 599  G74  G43 0.01126948 0.022642514
## 600  G80  G16 0.01126417 0.030648737
## 601  G57  G51 0.01125322 0.046957801
## 602  G42  G61 0.01122055 0.053521878
## 603  G44   G7 0.01120900 0.013840058
## 604  G56  G92 0.01120039 0.009130645
## 605  G35  G72 0.01119410 0.010141225
## 606  G76  G78 0.01118380 0.052332750
## 607  G50 G100 0.01118351 0.018616532
## 608  G82   G3 0.01115564 0.010954743
## 609  G47  G40 0.01113746 0.019852665
## 610  G44  G85 0.01113050 0.012978914
## 611  G79  G18 0.01112511 0.009838404
## 612  G31  G71 0.01111351 0.033590976
## 613  G25  G68 0.01108526 0.009981104
## 614   G8  G86 0.01107934 0.024254381
## 615  G53  G19 0.01104490 0.015081495
## 616  G52  G54 0.01103989 0.209010834
## 617  G76  G28 0.01102084 0.219401489
## 618  G13  G22 0.01101281 0.020753609
## 619  G55  G95 0.01101129 0.017296392
## 620  G82  G61 0.01099485 0.009630219
## 621  G81  G79 0.01096874 0.014678699
## 622  G85  G44 0.01095537 0.018304924
## 623   G5  G90 0.01094671 0.011904048
## 624  G68  G77 0.01091604 0.015218679
## 625  G55  G62 0.01091348 0.014241262
## 626  G27  G16 0.01091042 0.038545102
## 627  G12   G6 0.01090995 0.044843712
## 628  G27  G33 0.01090853 0.011222186
## 629  G76  G73 0.01090572 0.017546840
## 630  G90 G100 0.01087339 0.011563191
## 631  G69   G6 0.01087143 0.083811480
## 632  G33  G17 0.01083549 0.012155122
## 633  G13  G98 0.01082936 0.009583639
## 634  G51   G9 0.01082820 0.009158000
## 635  G26  G10 0.01080367 0.015016506
```

```
## 636   G4  G75 0.01079443 0.022248927
## 637  G58  G27 0.01078080 0.013631046
## 638  G38  G27 0.01077737 0.010191054
## 639  G21  G88 0.01076828 0.119563008
## 640  G55  G26 0.01076706 0.010049733
## 641  G80  G26 0.01076270 0.009355551
## 642  G32   G1 0.01075956 0.102490372
## 643  G22   G4 0.01075588 0.009092789
```

Q3) We have MC $X \rightarrow Y \rightarrow Z$

Where $X \perp Z \mid Y$ then $I(X,Y) \geq I(X,Z)$ [to Prove]

We know, $H(X,Y) \leq H(X) + H(Y)$

$\Rightarrow H((X, Y)|Z) \leq H(X|Z) + H(Y|Z)$

$\Rightarrow H(X|Z) \geq H((X,Y)|Z) - H(Y|Z)$

$\Rightarrow H(X|Z) \geq H(X, Y|Z)$ $[\because H(Y|Z) \geq 0]$

$\Rightarrow H(X|Z) \geq H(X|Y)$ $[\because X \perp Z \mid Y \Rightarrow H(X, Y|Z) = H(X|Y)$

$\Rightarrow H(X) - H(X|Z) \leq H(X) - H(X|Y)$

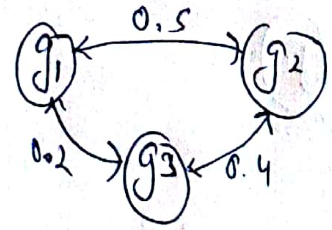
or $I(X,Z) \leq I(X,Y)$

3b) Data processing inequality (DPI): if we have $A \leftarrow B \leftrightarrow C$

then $I(A,C) \leq \min[I(A,B) ; I(B,C)]$

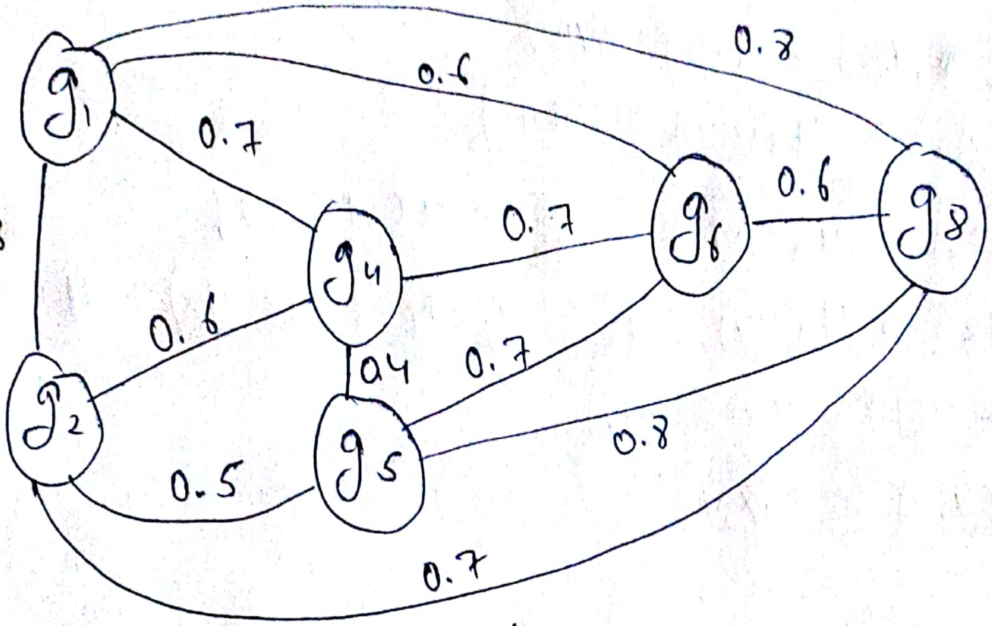
Using DPI we can remove links with smallest I in triangles.

e.g.

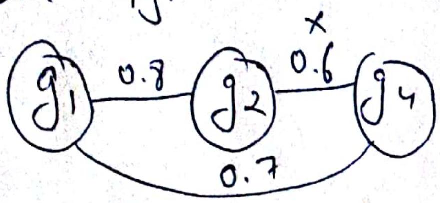


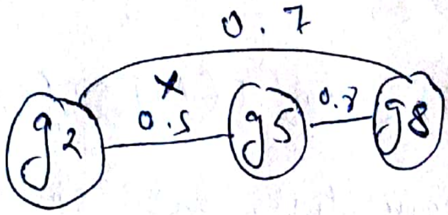
Using DPI we can remove link b/w g_1 and g_3 and simplify to $g_1 \overset{0.5}{\longleftrightarrow} g_2 \overset{0.4}{\longleftarrow} g_3$. This way we can simplify our network by removing low information edges.

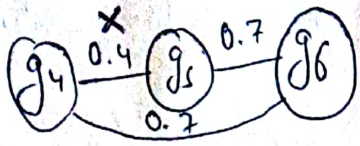
3c)

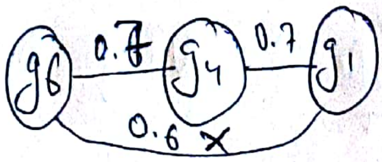


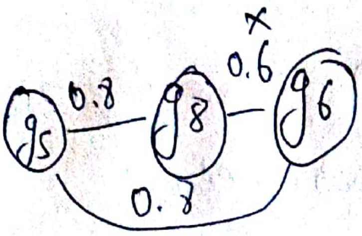
Starting at g_1 we find triangles and ~~remove~~ remove edges with lowest ~~ent~~ entropy.

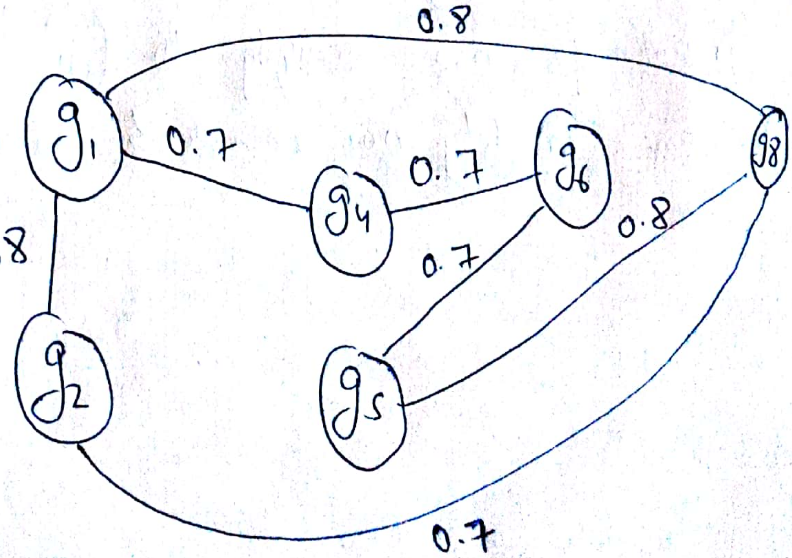












After Simplification