Validação

José Elvano Moraes

4/15/2021

Variáveis selecionadas

```
## Rows: 95,537
## Columns: 17
## $ IDADE      <fct> "(75,112]", "[1,38]", "(38,75]", "(38,75]", "(38,75]", "(38~
## $ FEBRE      <fct> 1, 1, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 1, 1, 1, 2, 1, 1, 2, 1,~
## $ GARGANTA   <fct> 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 2,~
## $ DISPNEIA   <fct> 2, 1, 1, 1, 2, 2, 2, 1, 1, 1, 2, 2, 2, 1, 2, 1, 2, 1, 1, 1,~
## $ SATURACAO  <fct> 1, 1, 1, 1, 1, 2, 2, 1, 1, 2, 1, 2, 2, 1, 2, 1, 1, 2, 2, 1,~
## $ EVOLUCAO   <fct> 1, 1, 2, 1, 1, 1, 1, 1, 3, 1, 1, 2, 1, 1, 2, 2, 2, 1, 2, 1,~
## $ RENAL      <fct> 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 1, 2, 2, 2, 2, 2, 1, 2,~
## $ DIABETES   <fct> 1, 2, 2, 2, 1, 1, 2, 2, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 1, 1,~
## $ OBESIDADE  <fct> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
## $ PERD_OLFT  <fct> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
## $ PERD_PALA  <fct> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
## $ NEUROLOGIC <fct> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
## $ PNEUMOPATI <fct> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
## $ UTI        <fct> 2, 2, 1, 2, 1, 1, 2, 2, 1, 1, 1, 2, 2, 2, 1, 2, 2, 2, 2, 2,~
## $ CARDIOPATI <fct> 1, 2, 2, 2, 2, 1, 2, 1, 2, 1, 2, 1, 2, 2, 2, 1, 1, 1, 1, 1,~
## $ SUPORT_VEN <fct> 2, 2, 2, 2, 2, 3, 2, 1, 1, 9, 2, 3, 3, 2, 2, 1, 2, 3, 2, 2,~
## $ ANTIVIRAL  <fct> 2, 2, 2, 1, 2, 2, 2, 9, 2, 9, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
```

Imposição de estrutura com arcos que fazem sentido clínico (*White list*)

```
s1 <- sample_frac(ddf, .8, FALSE)
cat(nrow(ddf) - nrow(s1))

## 19107

wl = matrix(c("OBESIDADE", "DIABETES",
              "IDADE", "DIABETES",
              "SATURACAO", "SUPORT_VEN",
              "IDADE", "SUPORT_VEN",
              "IDADE", "NEUROLOGIC",
              "DIABETES", "RENAL",
              "DIABETES", "CARDIOPATI",
              "UTI", "EVOLUCAO"),
       ncol = 2, byrow = TRUE, dimnames = list(NULL, c("from", "to")))

bn1 <- mmhc(s1, whitelist = wl)
```
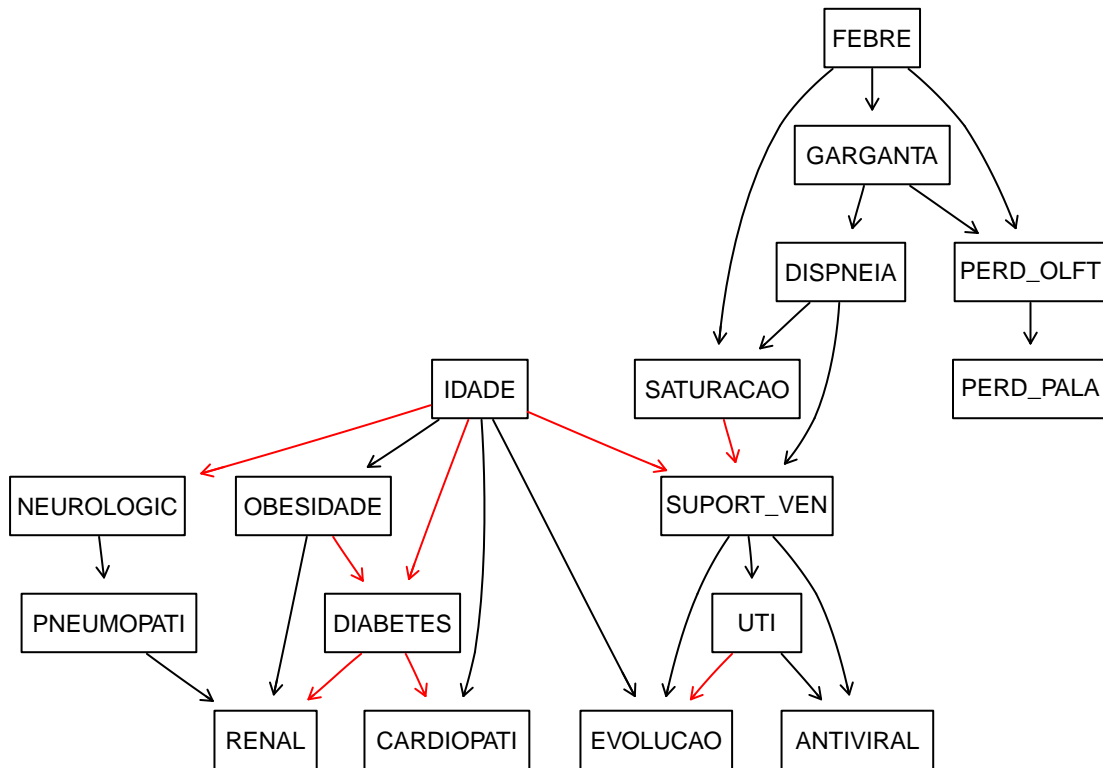
Rede Causal (DAG, *Directed Acyclic Graphic*)

```
#par(mfrow=c(2,2))
graphviz.plot(bn1, shape='rectangle', highlight = list(arcs = wl), main = '...')
```

...


```
fitted.1 <- bn.fit(bn1, s1)
```

```
#par(mfrow=c(2,2))
graphviz.chart(fitted.1, scale = c(2, 3),  type = "barprob", col = "darkblue", bg = "azure", bar.col =
```

```
## Loading required namespace: gRain

##
## Attaching package: 'gRbase'

## The following objects are masked from 'package:bnstruct':
##
##     dag, observations, observations<-

## The following objects are masked from 'package:igraph':
##
##     is_dag, topo_sort

## The following objects are masked from 'package:bnlearn':
##
##     ancestors, children, parents
```
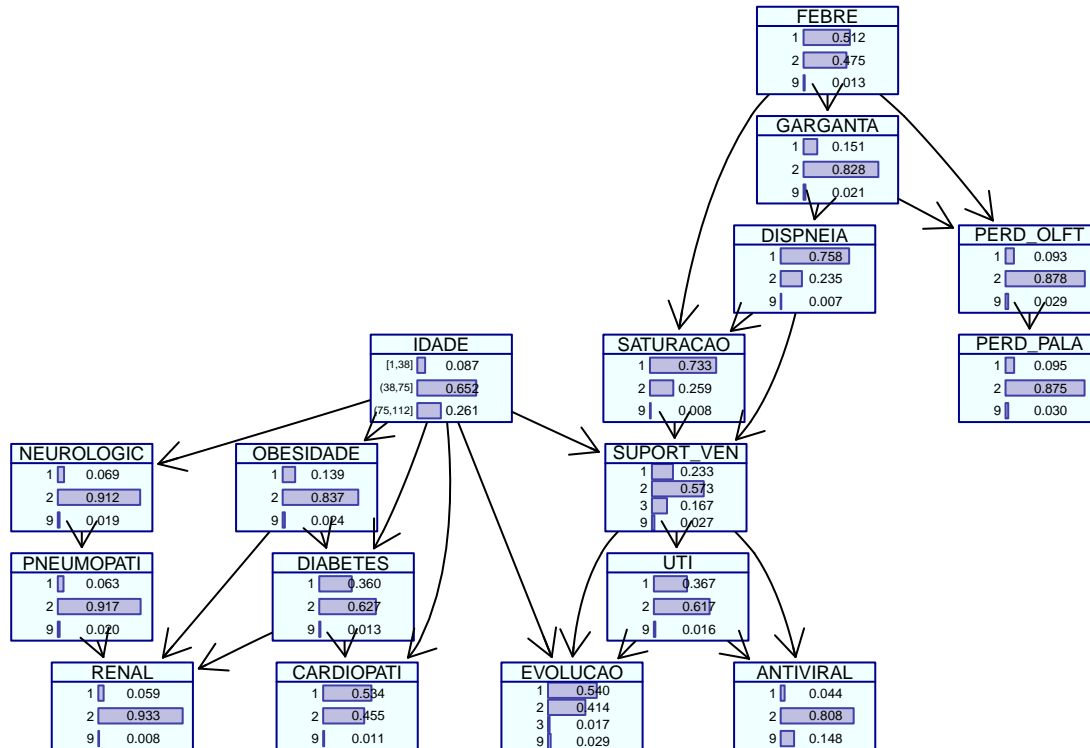
Rede de probabilidades

FEBRE
1	0.512
2	0.475
9	0.013

GARGANTA
1	0.151
2	0.828
9	0.021

DISPNEIA
1	0.758
2	0.235
9	0.007

PERD_OLFT
1	0.093
2	0.878
9	0.029

PERD_PALA
1	0.095
2	0.875
9	0.030

SATURACAO
1	0.733
2	0.259
9	0.008

IDADE
[1,38]	0.087
(38,75]	0.652
(75,112]	0.261

NEUROLOGIC
1	0.069
2	0.912
9	0.019

OBESIDADE
1	0.139
2	0.837
9	0.024

SUPORT_VEN
1	0.233
2	0.573
3	0.167
9	0.027

PNEUMOPATI
1	0.063
2	0.917
9	0.020

DIABETES
1	0.360
2	0.627
9	0.013

UTI
1	0.367
2	0.617
9	0.016

RENAL
1	0.059
2	0.933
9	0.008

CARDIOPATI
1	0.534
2	0.455
9	0.011

EVOLUCAO
1	0.540
2	0.414
3	0.017
9	0.029

ANTIVIRAL
1	0.044
2	0.808
9	0.148

Bootstrapping

```
boots.trap  <- 400
str.diff = suppressMessages(boot.strength(s1, R = boots.trap, algorithm = "mmhc"))
cat(paste('Threshold: ', attr(str.diff, "threshold")))
```
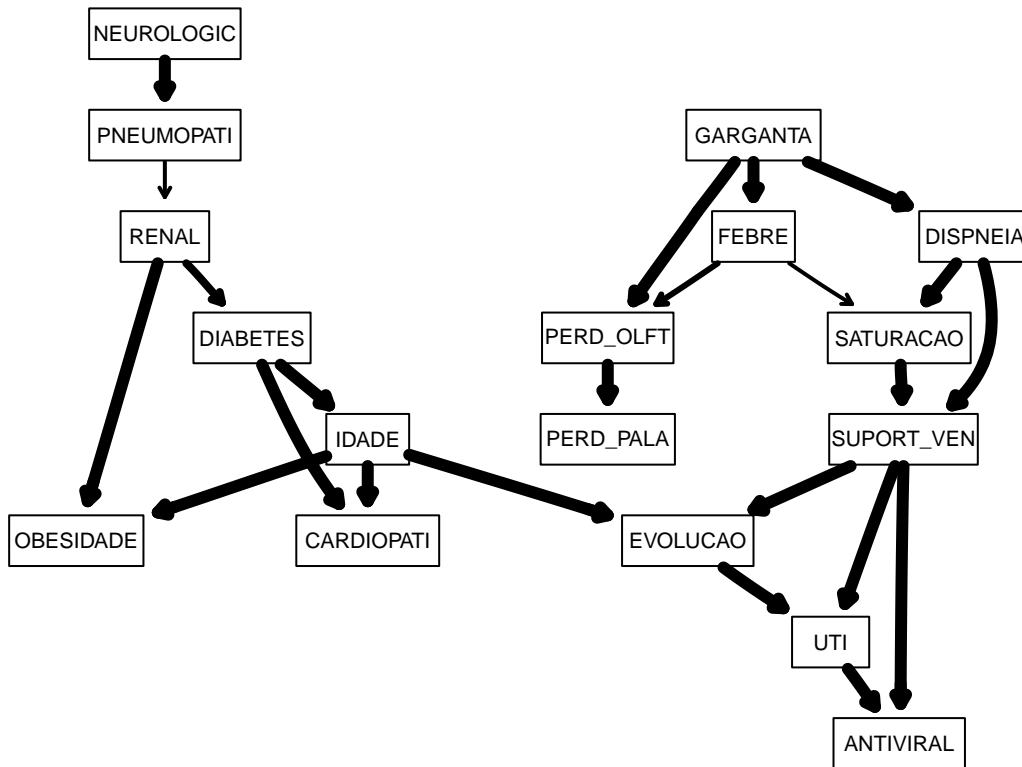
```
## Threshold:  0.445
```

```
avg.diff = averaged.network(str.diff)
```

```
## Warning in averaged.network.backend(strength = strength, nodes = nodes, : arc
## IDADE -> NEUROLOGIC would introduce cycles in the graph, ignoring.
```

```
strength.plot(avg.diff, str.diff, shape = "rectangle", main = paste("Iterações = ", boots.trap))
```

Iterações = 400

How can we compare the averaged network (avg.diff) with the network we originally learned in from all the data? The most qualitative way is to plot the two networks side by side, with the nodes in the same positions, and highlight the arcs that appear in one network and not in the other, or that appear with different directions.

```
par(mfrow = c(1, 2))
graphviz.compare(avg.diff, bn1, shape = "rectangle", main = c("DAG médio", "DAG único"))
```
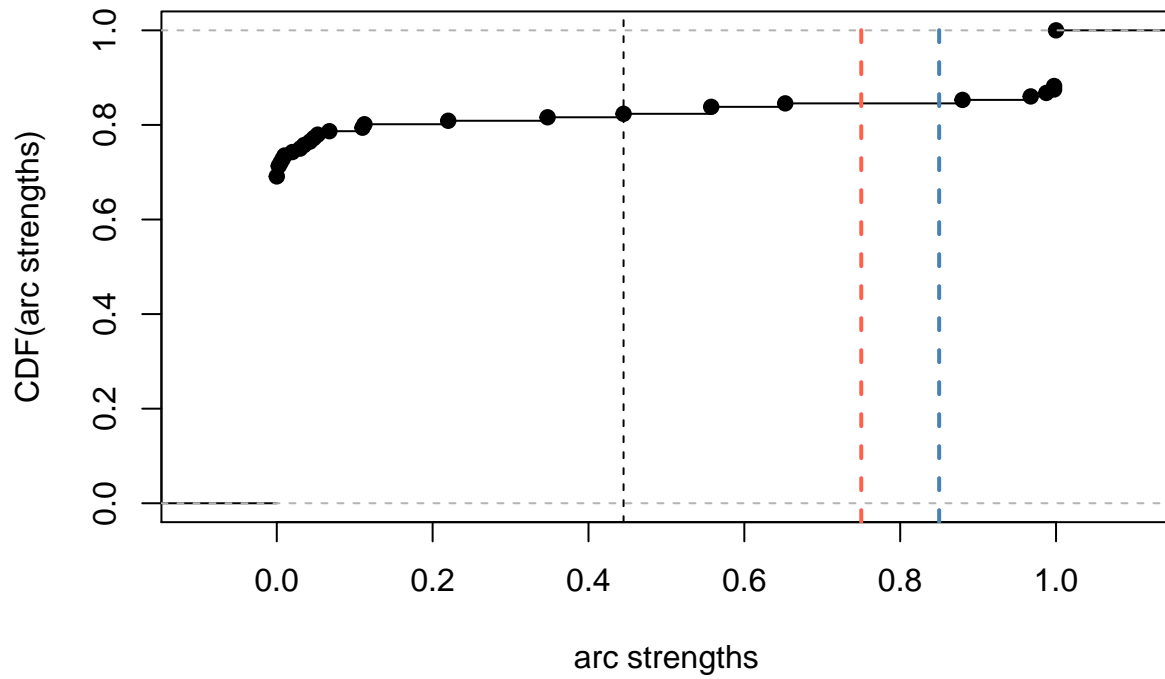
DAG médio
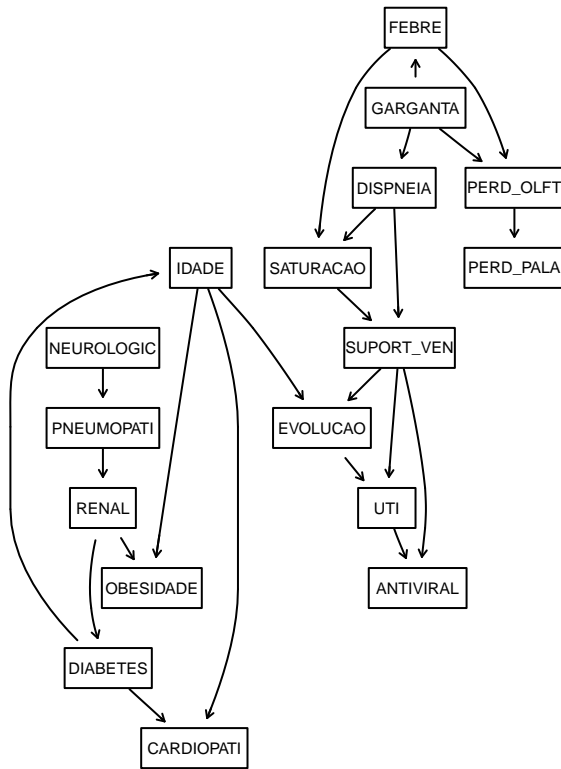
DAG único
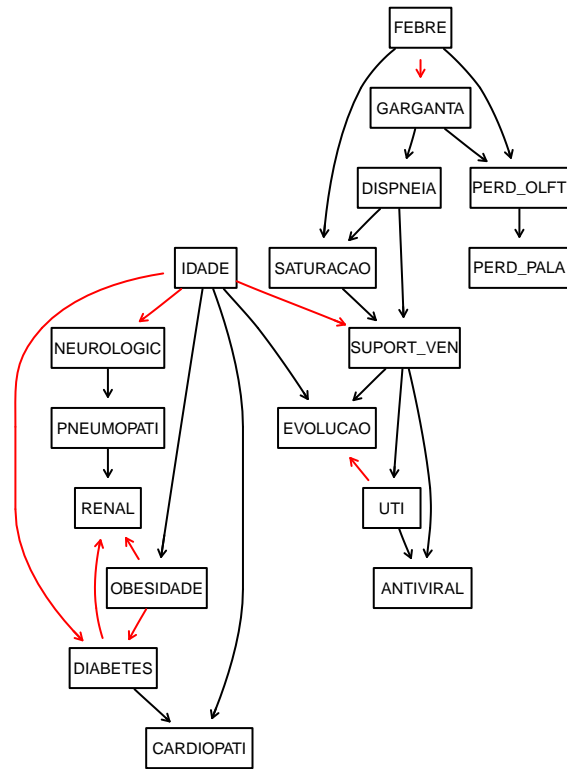
It is also a good idea to look at the threshold with respect to the distribution of the arc strengths

```
plot(str.diff)
abline(v = 0.75, col = "tomato", lty = 2, lwd = 2)
abline(v = 0.85, col = "steelblue", lty = 2, lwd = 2)
```
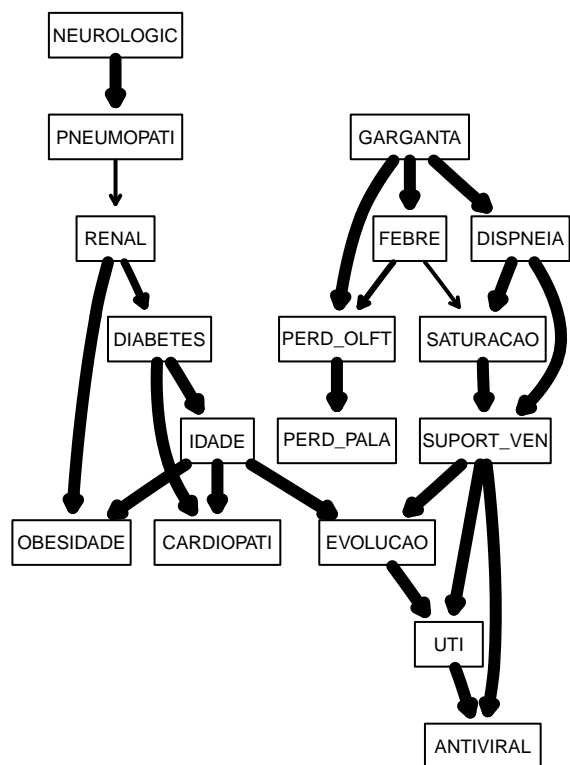
threshold = 0.445

arc strengths

The simpler network we obtain by setting threshold = 0.8 in averaged.network() is shown below; it is certainly easier to reason with from a qualitative point of view.

```
par(mfrow = c(1, 2))
avg.simpler = averaged.network(str.diff, threshold = 0.75)
strength.plot(avg.diff, str.diff, shape = "rectangle", main = paste("Iterações = ", boots.trap, " Thr =
strength.plot(avg.simpler, str.diff, shape = "rectangle", main = 'Iterações = 100 Thr = 0.75')
```

Iterações = 400 Thr = 0.445

Iterações = 100 Thr = 0.75