

## *Belief Propagation over a Bayesian Network*

---

**Belief Propagation:** If there is a variable X in a cluster and it carries its own belief of the variable X and even it applies to other clusters having X. There exists a unique path between such clusters. If X is changed in one of the clusters we can verify how it will affect in other clusters as there exists a path and each cluster starts updating the new belief of X and act accordingly and we can verify the new probability of the cluster and determines predictions.

**Packages Used:** gRain, igraph, ggm

**Dataset:** Coronary artery disease data

**Methods:** dSep, cptable, grain, compileCPT, extractCPT, setFinding, getFinding, querygrain, simulate.grain

**Goal:** Construct the network, identify nodes that are d-separable, structure learning and parameter learning and then belief propagation and then simulation of the model over a data to verify results and to make predictions.

**Newly created functions:**

**ffinddsepingraph-** function which returns all possible d-separations in the given input graph

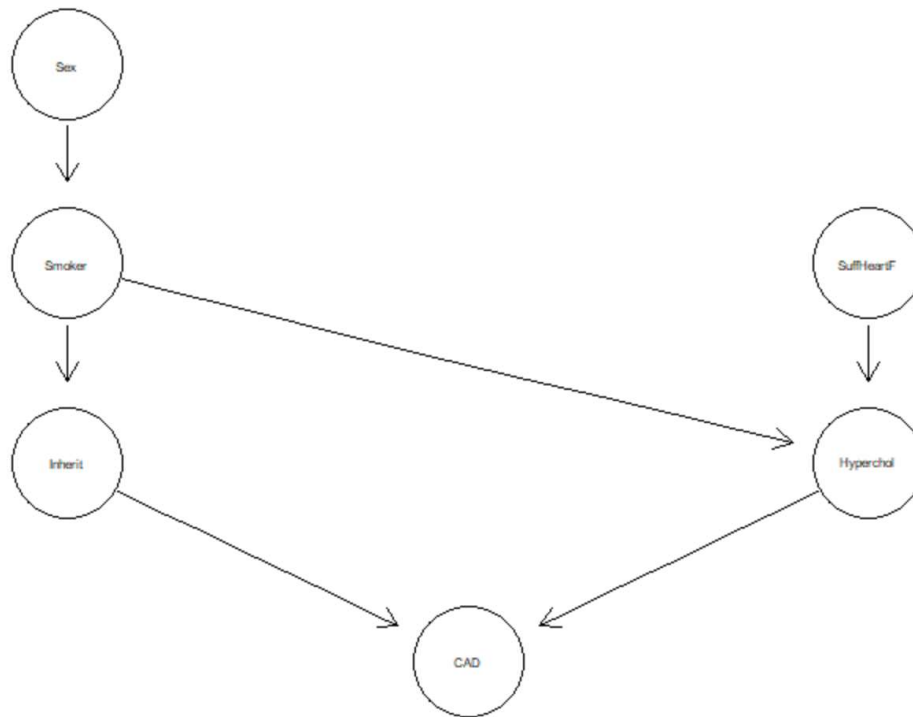
CAD dataset contains 14 columns but we will use just 6 columns to create a network and check belief propagation

```
> #load data cad1 Coronary artery disease data
> data(cad1)
> ?cad1
>
> #236 records and 14 variables
> dim(cad1)
[1] 236 14
> names(cad1)
[1] "Sex" "AngPec" "AMI" "QWave" "QWavecode" "STcode" "STchange"
[8] "SuffHeartF" "Hypertroph" "Hypercho" "Smoker" "Inherit" "Heartfail" "CAD"
```

**Reduced dataset with required columns:**

```
> cad2<-cad1[,c(1,8,10,11,12,14)]
> names(cad2)
[1] "Sex" "SuffHeartF" "Hypercho" "Smoker" "Inherit" "CAD"
> head(cad2)
  Sex SuffHeartF Hypercho Smoker Inherit CAD
1  Male         NO         NO     NO     NO  NO
2  Male         NO         NO     NO     NO  NO
3 Female         NO         NO     NO     NO  NO
4  Male         NO         NO     NO     NO  NO
5  Male         NO         NO     NO     NO  NO
6  Male         NO         NO     NO     NO  NO
> dim(cad2)
[1] 236 6
```

## 1. Constructing a network



## 2. Finding all possible d-separations in the graph – total of 33

```

> cdagmat<-as(cdag,"matrix")
> ffinddsepingraph(cdagmat)
Nodes Sex and SuffHeartF are d-separated given NULL
Nodes Sex and SuffHeartF are d-separated given Smoker
Nodes Sex and SuffHeartF are d-separated given Inherit
Nodes Sex and SuffHeartF are d-separated given Smoker Inherit
Nodes Sex and SuffHeartF are d-separated given Smoker Inherit Hyperchol
Nodes Sex and SuffHeartF are d-separated given Smoker Inherit Hyperchol CAD
Nodes Sex and Inherit are d-separated given Smoker
Nodes Sex and Inherit are d-separated given Smoker SuffHeartF
Nodes Sex and Inherit are d-separated given Smoker SuffHeartF Hyperchol
Nodes Sex and Inherit are d-separated given Smoker SuffHeartF Hyperchol CAD
Nodes Sex and Hyperchol are d-separated given Smoker
Nodes Sex and Hyperchol are d-separated given Smoker SuffHeartF
Nodes Sex and Hyperchol are d-separated given Smoker SuffHeartF Inherit
Nodes Sex and Hyperchol are d-separated given Smoker SuffHeartF Inherit CAD
Nodes Sex and CAD are d-separated given Smoker
Nodes Sex and CAD are d-separated given Smoker SuffHeartF
Nodes Sex and CAD are d-separated given Smoker SuffHeartF Inherit
Nodes Sex and CAD are d-separated given Smoker SuffHeartF Inherit Hyperchol
Nodes Smoker and SuffHeartF are d-separated given NULL
Nodes Smoker and SuffHeartF are d-separated given Sex
Nodes Smoker and SuffHeartF are d-separated given Inherit
Nodes Smoker and SuffHeartF are d-separated given Sex Inherit
Nodes Smoker and CAD are d-separated given Sex SuffHeartF Inherit Hyperchol
Nodes SuffHeartF and Inherit are d-separated given NULL
Nodes SuffHeartF and Inherit are d-separated given Sex
Nodes SuffHeartF and Inherit are d-separated given Smoker
Nodes SuffHeartF and Inherit are d-separated given Sex Smoker
Nodes SuffHeartF and Inherit are d-separated given Sex Smoker Hyperchol
Nodes SuffHeartF and Inherit are d-separated given Sex Smoker Hyperchol CAD
Nodes SuffHeartF and CAD are d-separated given Sex Smoker Inherit Hyperchol
Nodes Inherit and Hyperchol are d-separated given Smoker
Nodes Inherit and Hyperchol are d-separated given Sex Smoker
Nodes Inherit and Hyperchol are d-separated given Sex Smoker SuffHeartF

```

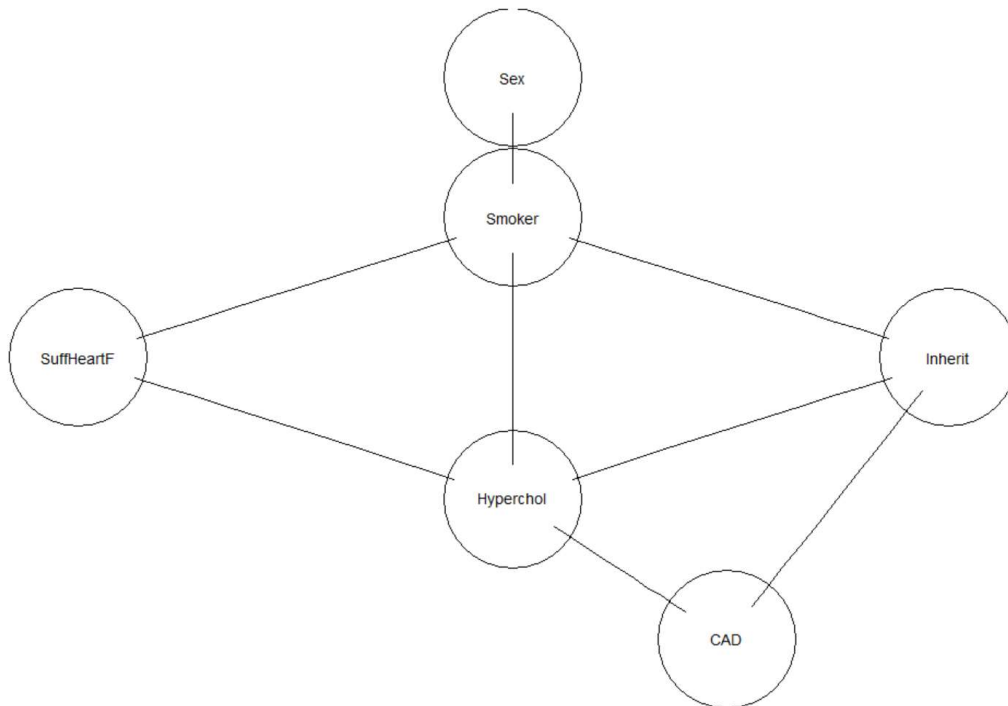
### 3. Setting CP Tables:

```
> #To Get conditional probability tables
> yn<-c("Yes","No")
> mf<-c("Male","Female")
> sexcpt<- cptable(~Sex,values=c(0.8,0.2),levels = mf)
> smok.sex<- cptable(~Smoker|Sex,values =c(0.6,0.4,0.99,0.01),levels = yn)
> suffh<-cptable(~SuffHeartF,values = c(0.6,0.4),levels = yn)
> inh.smok<-cptable(~Inherit|Smoker,values = c(0.8,0.2,0.6,0.4),levels = yn)
> hyp.smok.suff<-cptable(~Hyperchol|Smoker:SuffHeartF,values = c(0.5,0.5,0.7,0.3,0.9,0.1,0.4,0.6),levels = yn)
> cad.inh.hyp<-cptable(~CAD|Inherit:Hyperchol,values = c(0.6,0.4,0.5,0.5,0.8,0.2,0.4,0.6),levels = yn)
>
```

### 4. Build the network:

```
> plist<- compileCPT(list(sexcpt, smok.sex, suffh, inh.smok, hyp.smok.suff, cad.inh.hyp))
> grn1<- grain(plist)
> summary(grn1)
Independence network: Compiled: TRUE Propagated: FALSE
Nodes : chr [1:6] "Sex" "Smoker" "SuffHeartF" "Inherit" "Hyperchol" "CAD"
Number of cliques:      4
Maximal clique size:    3
Maximal state space in cliques: 8
```

### 5. Network built



## 6. CPT table details

```
> cptdata
$Sex
Sex
  Female      Male
0.1991525 0.8008475

$Smoker
  Sex
Smoker  Female      Male
  No    0.3617021 0.1798942
  Yes   0.6382979 0.8201058

$SuffHeartF
SuffHeartF
  No      Yes
0.7076271 0.2923729

$Inherit
  Smoker
Inherit  No      Yes
  No     0.8235294 0.6486486
  Yes    0.1764706 0.3513514

$Hyperchol
, , SuffHeartF = No
  Smoker
Hyperchol  No      Yes
  No       0.6750000 0.4645669
  Yes      0.3250000 0.5354331

, , SuffHeartF = Yes
  Smoker
Hyperchol  No      Yes
  No       0.2727273 0.3275862
  Yes      0.7272727 0.6724138

$CAD
, , Hyperchol = No
  Inherit
CAD      No      Yes
  No     0.8214286 0.5000000
  Yes    0.1785714 0.5000000

, , Hyperchol = Yes
  Inherit
CAD      No      Yes
  No     0.4487179 0.2600000
  Yes    0.5512821 0.7400000

attr("graph")
A graphNEL graph with directed edges
Number of Nodes = 6
Number of Edges = 6
attr("class")
[1] "cpt_rep"
```

## 7. Compile and propagate the network

```
> grn1c<- propagate(grn1c)
> summary(grn1c)
Independence network: Compiled: TRUE Propagated: TRUE
Nodes : chr [1:6] "Sex" "Smoker" "SuffHeartF" "Inherit" "Hyperchol" "CAD"
Number of cliques:      4
Maximal clique size:    3
Maximal state space in cliques: 8
```

## 8. Setting new finding onto network

```
> grn1c.ev<- setFinding(grn1c,nodes=c("Sex","Hyperchol"), states=c("Female","Yes"))
> #Probability of finding
> getFinding(grn1c.ev)
      nodes is.hard.evidence hard.state
1      Sex              TRUE      Female
2 Hyperchol              TRUE         Yes
```

## 9. Query to verify the absorbed evidence before and after

```
> #Not absorbed - Before evidence
> querygrain(grn1c,nodes = c("CAD","SuffHeartF"),type="marginal")
$SuffHeartF
SuffHeartF
Yes  No
0.6  0.4

$CAD
CAD
      Yes      No
0.616652 0.383348

>
> #Absorbed - After evidence
> querygrain(grn1c.ev,nodes = c("CAD","SuffHeartF"),type="marginal")
$SuffHeartF
SuffHeartF
      Yes      No
0.4569175 0.5430825

$CAD
CAD
      Yes      No
0.579824 0.420176
```

After the belief propagation, it shows that the female with high cholesterol has chances of heart failure of 45% and CAD chances of 57.9%

## 10. Simulation for 25 observations

```
> sim25<-simulate.grain(grn1c.ev,nsim=25,seed=100)
> table(sim25$SuffHeartF)

Yes  No
  9   16
> table(sim25$CAD)

Yes  No
 12   13
>
> Heartfailurefreq<-table(sim25$SuffHeartF)
> probHfail25<-Heartfailurefreq[1]/length(sim25$SuffHeartF)
> probHfail25
Yes
0.36
>
> CADfreq<-table(sim25$CAD)
> probCAD25<-CADfreq[1]/length(sim25$CAD)
> probCAD25
Yes
0.48
```

When simulated 25 observations over a propagated belief, it shows the out of 25 females 9 might have chances of heart failure and 12 with CAD who has high cholesterol

**For 25 observations:**

**Probability of heart failure – 36%**

**Probability of CAD - 48%**

## 11. Simulation for 500 observations

```
> sim500<-simulate.grain(grn1c.ev,nsim=500,seed=100)
> table(sim500$SuffHeartF)

Yes  No
231  269
> table(sim500$CAD)

Yes  No
277  223
>
> Heartfailurefreq<-table(sim500$SuffHeartF)
> probHfail500<-Heartfailurefreq[1]/length(sim500$SuffHeartF)
> probHfail500
Yes
0.462
>
> CADfreq<-table(sim500$CAD)
> probCAD500<-CADfreq[1]/length(sim500$CAD)
> probCAD500
Yes
0.554
```

Out of 500 females with high cholesterol, 231 might get heart failure and 277 with CAD.

**Outcome of 500 observations:**

Probability of Heart Failure – 46%

Probability of CAD - 55%

**Observation:** Predicted data for 25 observations and 500 observations is nearly same at least for CAD. With graphical modelling, structure and parameter learning and belief propagation we can predict the results which is nearly to reality. Predicted results can be put into use to device action plan for public safety measures or raising awareness of high cholesterol levels in females and national nutrition can work out dietary plan.