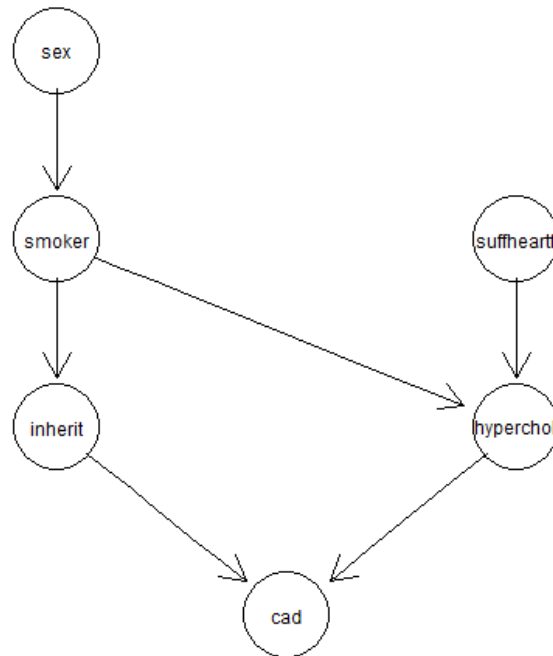


Sai Kiran Putta

1)

a)

Building the network and plotting it gives the following plot.



b)

Joint probabilities before and after absorbing the evidence is given below.

```
> querygrain(grain1_compile, nodes = c("SuffHeartF", "CAD"), type = "joint")
  SuffHeartF
CAD      No      Yes
No  0.3957368 0.1443930
Yes 0.3118903 0.1479799
attr(,"class")
[1] "parray" "array"
> querygrain(grain1_compile_ev, nodes = c("SuffHeartF", "CAD"), type = "joint")
  SuffHeartF
CAD      No      Yes
No  0.4078210 0.1453059
Yes 0.2998061 0.1470670
attr(,"class")
```

Conditional probabilities before and after absorbing the evidence is given below.

```
> querygrain(grain1_compile, nodes = c("SuffHeartF", "CAD"), type = "conditional")
SuffHeartF
CAD      No      Yes
No  0.7326698 0.2673302
Yes 0.6782138 0.3217862
attr(,"class")
[1] "parray" "array"
> querygrain(grain1_compile_ev, nodes = c("SuffHeartF", "CAD"), type = "conditional")
SuffHeartF
CAD      No      Yes
No  0.7373010 0.2626990
Yes 0.6708976 0.3291024
attr(,"class")
[1] "parray" "array"
```

Marginal probabilities before and after absorbing the evidence is given below.

```
> querygrain(grain1_compile, nodes = c("SuffHeartF", "CAD"), type = "marginal")
$CAD
CAD
      No      Yes
0.5401298 0.4598702

$SuffHeartF
SuffHeartF
      No      Yes
0.7076271 0.2923729

> querygrain(grain1_compile_ev, nodes = c("SuffHeartF", "CAD"), type = "marginal")
$CAD
CAD
      No      Yes
0.5531269 0.4468731

$SuffHeartF
SuffHeartF
      No      Yes
0.7076271 0.2923729
```

C)

Following are the 5 new data points that are generated.

```
> sim_find5
      Sex Smoker Inherit CAD Hyperchol SuffHeartF
1 Female    No     No   No      No        No
2 Female   Yes     Yes  Yes     Yes        No
3 Female   Yes     Yes  Yes     Yes        No
4 Female   Yes     Yes  No      No        No
5 Female   Yes     No   No      Yes        Yes
> |
```

Following are the predictions of the newly generated 5 new data points.

```
$pred
$pred$Smoker
[1] "No" "Yes" "Yes" "Yes" "Yes"

$pred$CAD
[1] "No" "Yes" "Yes" "No" "Yes"

$Evidence
[1] 0.05544145 0.01984580 0.01984580 0.02075442 0.02882428
```

d) After generating new 500 data points and calculating the mis-classification rate, following are the results.

Mis-classification rate for smoker is 34.2 %

Mis-classification rate for CAD is 32.4 %

Since the performance of the network/classifier depends on the data, we cannot say if the accuracy is good enough. But the classifier doing ~15% better than a random guess.

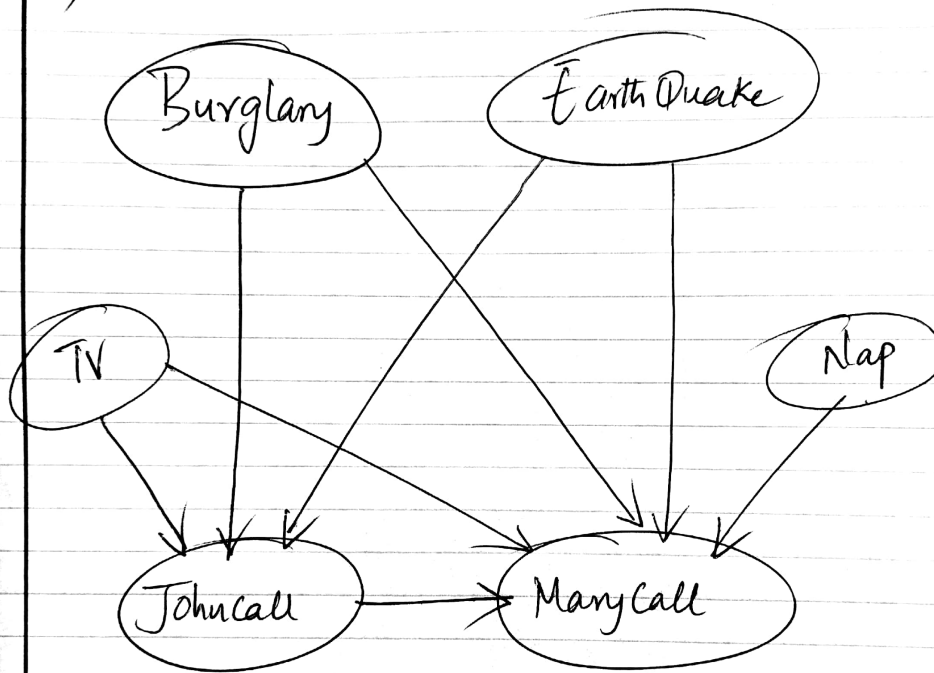
There might be other “valid combinations” of positions of nodes in the network which might lead to better accuracy.

Different versions could be tried to arrive at best performing network.

2)

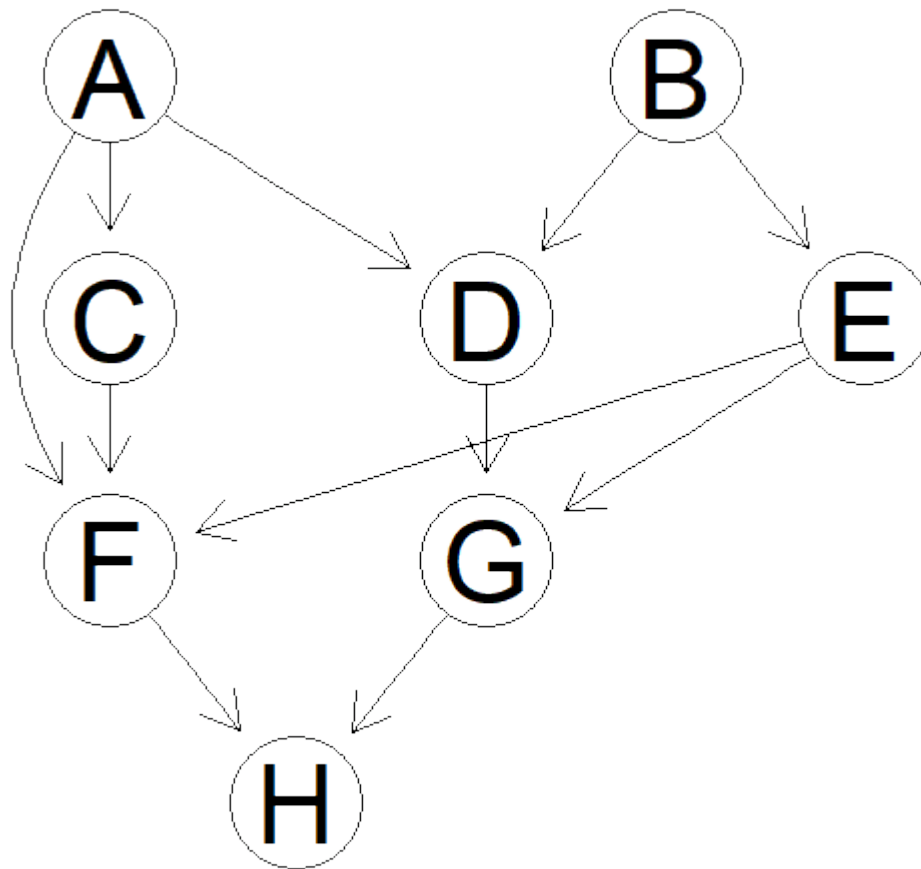
Sai Kiran putta.

2)



3)

Following is the network that is obtained.



Following are the results that were obtained for the questions.

A) C and G are d-separated. – ‘FALSE’

B) C and E are d-separated. – ‘TRUE’

C) C and E are d-connected given evidence about G. – ‘FALSE’

D) A and G are d-connected given evidence about D and E. – ‘TRUE’

E) A and G are d-connected given evidence on D. – ‘FALSE’