Homework for lecture 11&12

**Instructor**: Pei Wang

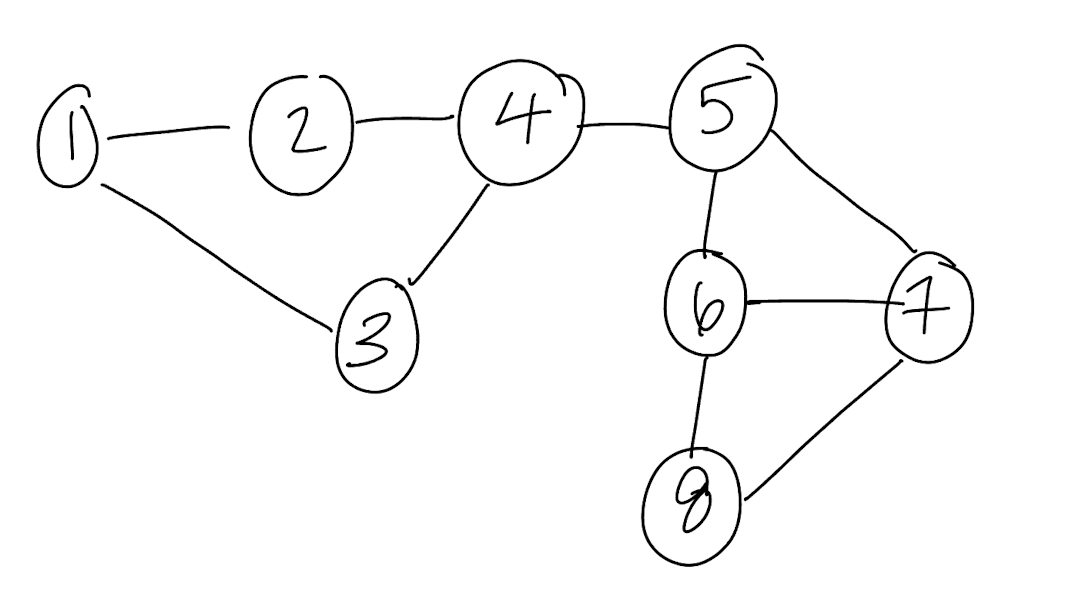
Modified 04/10/23 by Li Shen

*For the programming exercises, also include the code used in your report.*

1. Consider a “medical diagnostic” graph as illustrated in below. 

**[15 points]** Find at least 4 conditional independence relationships according to the graph.

1. Muscle pain is conditionally independent of congestion.
2. Flu is conditionally independent of hay fever.
3. Congestion is conditionally independent from vaccine.
4. Congestion is conditionally independent from season.
5. Please follow the below instructions.
   1. **[5 points]** Manually draw the undirected graph with 8 nodes and edges between the following pairs of nodes:  
      (1, 2) (2,4) (1, 3) (3,4) (4,5) (5,6) (6,7) (6,8) (5,7) (7,8)

****

* 1. A picture containing text

     Description automatically generated**[5 points]** Write out the degree matrix of this graph.
  2. **[5 points]** Calculate the edge-betweenness score for edge (4,5).

Edge-betweenness = 4x8/2 = 16 \*\*\*see attached calculations at the end of document

* 1. **[5 points]** Calculate the degree centrality scores for node 4, 5, 6.

Degree centrality node 4 = 3 (standardized score = 3/n\_nodes-1 or 3/7)

Degree centrality node 5 = 3 (standardized score = 3/n\_nodes-1 or 3/7)

Degree centrality node 6 = 3 (standardized score = 3/n\_nodes-1 or 3/7)

* 1. **[5 points]** Calculate the betweenness centrality scores for node 4,5 and 6.

Betweenness centrality score node 4 = 2+1+1+1+1+1+2 = 9 and standardized score is 9/ (n-1)(n-2)/2= 9/21=3/7

Betweenness centrality score node 5 = 2+1+1+1+1+1+2 = 9 and standardized score is 9/21 = 3/7

Betweenness centrality score node 6 = 1+1+1+1+1+1+2 = 8 and standardized score is 8/21

\*\*\*see attached calculations at the end of document

**[10 points]** Write an R script to create a graph object (see igraph package) in R for the above graph.

library(igraph)

graph = graph(c(1,2, 2,4, 1,3, 3,4, 4,5, 5,6, 6,7, 6,8, 5,7, 7,8))  
plot(graph)

Chart, scatter chart

Description automatically generated

* 1. **[10 points]** Derive 2 modules using cluster\_edge\_betweenness() in the igraph package.

modules <- cluster\_edge\_betweenness(graph)

module 1 = 1,2,3,4

module 2 = 5,6,7,8

* 1. **[5 points]** Draw the topology of this graph using plot.igraph(). Color one module in red and another one in green.

dendPlot(modules, mode="hclust",colbar=c('red','green'))

Chart, box and whisker chart

Description automatically generated

1. This problem uses the data in the attached file "network\_assignment\_data\_set.txt", which lists the values that 6 variables take in 200 different samples/observations. Answer the following problems using your preferred Bayesian network learning package, such as the bnlearn package in R.
   1. **[20 points]** Build a DAG-structured Bayesian network of the 6 variables using the hill climbing algorithm. Plot the inferred network. (Hint: use the “hc” function in the *bnlearn* package, or the equivalent in the package you’re using).

library(bnlearn)

library(igraph)  
library(dplyr)

data = read.delim("~/Documents/GraduateSchool/Classes/ML Class/network\_assignment\_data\_set.txt")  
data = select(data,-X)  
cols = names(data)  
data[,cols] = lapply(data[,cols],factor)

bn.hc = hc(data, score = "bic")  
bn.hc

##   
## Bayesian network learned via Score-based methods  
##   
## model:  
## [A][C][B|A][D|A:C][E|B][F|E]   
## nodes: 6   
## arcs: 5   
## undirected arcs: 0   
## directed arcs: 5   
## average markov blanket size: 2.00   
## average neighbourhood size: 1.67   
## average branching factor: 0.83   
##   
## learning algorithm: Hill-Climbing   
## score: BIC (disc.)   
## penalization coefficient: 2.649159   
## tests used in the learning procedure: 40   
## optimized: TRUE

plot(bn.hc, main = "Hill-Climbing")

Diagram

Description automatically generated

* 1. **[5 points]** Plot (or draw) the moral graph of the inferred DAG.

library(gRbase)

library(bnlearn)  
  
bn.hc.moral = moral(bn.hc)  
plot(bn.hc.moral)

Diagram

Description automatically generated

* 1. **[10 points]** Write out the conditional probability table of the edge E->F in the above DAG, i.e., the value of P(F=a|E=a) and all other such combinations. (Hint: use the “bn” function in the *bnlearn* package, or the equivalent in the package you’re using.)

fitted.bn.hc = bn.fit(bn.hc, data = data)  
fitted.bn.hc$F

##   
## Parameters of node F (multinomial distribution)  
##   
## Conditional probability table:  
##   
## E  
## F a b c  
## a 0.5232558 0.2096774 0.6730769  
## b 0.4767442 0.7903226 0.3269231

Edge-betweenness calculations for Question 2c:

Diagram

Description automatically generated

Diagram

Description automatically generated

Diagram

Description automatically generated

Citation for edge betweenness methods: <https://medium.com/analytics-vidhya/girvan-newman-the-clustering-technique-in-network-analysis-27fe6d665c92#:~:text=Select%20a%20node%20X%2C%20and,as%20score%20to%20each%20node.&text=Sum%20up%20all%20of%20the,the%20edge%20betweenness%20of%20edges.>