Statistical Data Mining II
Homework 3

Due: Wednesday April 6th (11:59 pm)
40 points

Directions: See "homework guidelines" on UB learns for detailed information.

1) (15 points) Consider two networks "Les Miserables" and "Dolphins". These networks can be accessed from the library "igraph", using the following:

```
> library(igraph)
Warning message:
package 'igraph' was built under R version 3.1.2
> nexus.get("karate")
IGRAPH UNW- 34 78 -- Zachary's karate club network
+ attr: name (g/c), Citation (g/c), Author (g/c), Faction (v/n),
  name (v/c), weight (e/n)
> ?nexus.get
starting httpd help server ... done
> nexus.get("miserables")
IGRAPH UNW- 77 254 -- Les Miserables coappearance network
+ attr: name (g/c), Citation (g/c), Author (g/c), URL (g/c),
  Coappearance (g/x), name (v/c), Description (v/c), weight (e/n)
> nexus.get("dolphins")
IGRAPH UN-- 62 159 -- Dophin social network
+ attr: Description (g/c), name (g/c), Author (g/c), Citation (g/c),
  URL (g/c), name (v/c)
```

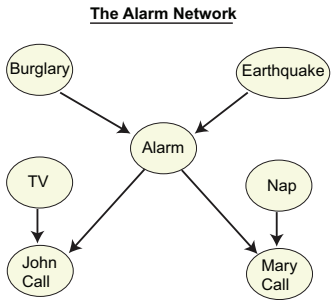
Using the hierarchical random graphs functions in "igraph" perform the following tasks:
(a) Find a consensus dendrogram that is based on MCMC-based sampling, and produce a plot that reveals communities.
(b) Focus on the dolphin network. Create noisy datasets. Do this by deleting 5% of the edges randomly (track which ones they are). Perform MCMC on this data followed by link-prediction. Are you able to predict the edges that you deleted at random well?
(c) Repeat the exercise in part (b) after deleting 15%, and 40%.

(Note: set.seed(1) before sampling, and see igraphdemo("hrg").

Citations:
Lusseau, D., Schneider, K., Boisseau, O.J., Haase, P., Slooten, E. & Dawson S.M. 2003. The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. Can geographic isolation explain this unique trait? Behavioral Ecology and Sociobiology 54(4): 396-405.

Les Miserables (Victor Hugo) Coappearance weighted network of characters in the novel Les Miserables.

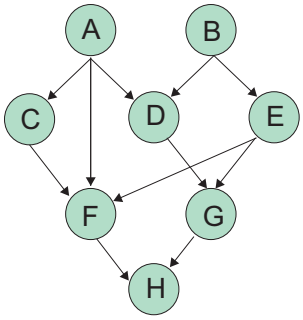
2) (10 points, adopted from exercise 3.11 in Koller et al.) Consider the following famous Bayesian Network by Judea Pearl.



The network is set up to answer questions of the following type:
"I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary does not call. Sometimes minor earthquakes set it off. Is there a burglar?"

One operation on Bayesian Networks that arises in many settings is the marginalization of some node in the network. Let the original Bayesian Network be denoted as B. Construct a Bayesian Network B' over all of the nodes EXCEPT for Alarm that is the minimal I-map for the marginal distribution $P_B(\mathrm{B}, \mathrm{E}, \mathrm{T}, \mathrm{N}, \mathrm{J}, \mathrm{M})$. Be sure to get all dependencies that remain from the original graph.

3) (10 points) Determine if the following statements are "TRUE OR FALSE" based on the DAG.



A) C and G are d-separated.
B) C and E are d-separated.
C) C and E are d-connected given evidence about G.
D) A and G are d-connected given evidence about D and E.
E) A and G are d-connected given evidence on D.