**STA 546 – Home Work #3**

**Suruchi JaiKumar Ahuja**

Q1). The library and R package igraph is used for network analysis. The main goals of the igraph library is to provide a set of data types and functions for pain-free implementation of graph algorithms, fast handling of large graphs, with millions of vertices and edges, allowing rapid prototyping via high level languages like R.

The network Les miserables (miserables) has 77 vertices and 254 edges. The network zachary’s karate club (karate) has 34 vertices and 78 edges. The network Dolphin social network (dolphins) has 62 vertices and 159 edges. These three networks were loaded into R environment using:

>> library(igraph)

>> nexus.get(“karate”)

>> nexus.get(“misearbles”)

>> nexus.get(“dolphins”)

1. Hrg model at level at level 5

Hierarchical random graph, at level 5:

g1 p=0.082 DN63

'- g44 p=0.006

'- g30 p= 0.23 Kringel

'- g60 p=0.094

'- g53 p=0.069 MN60 Five PL Beak SN96 Fish TR77 Bumper Cross

Oscar SMN5 Zipfel SN63 Whitetip TSN103 Grin Scabs Shmuddel

Stripes TR99 Hook Zap SN100 SN9 SN4 Thumper Double

Vau TR120 TSN83 TR88 CCL

'- g39 p= 1 MN105 Trigger MN83 Topless Patchback Haecksel Jonah

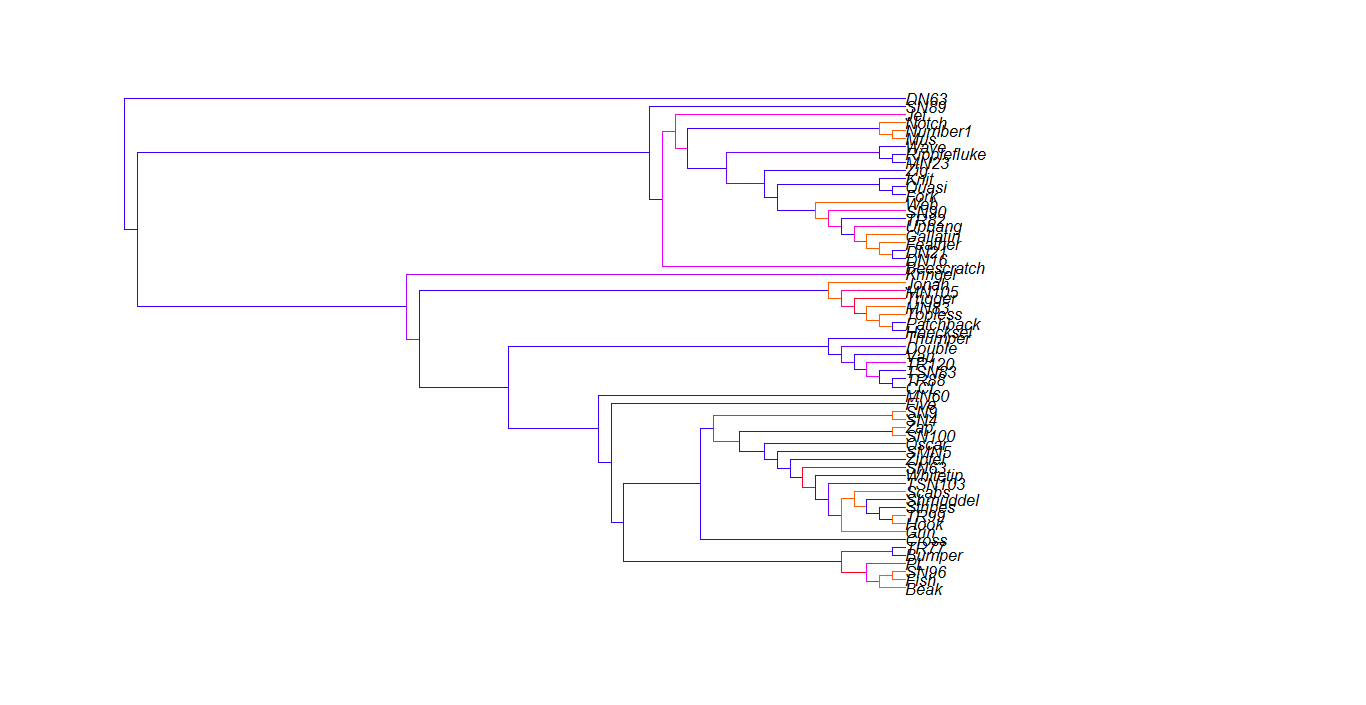
'- g41 p= 0.05 SN89

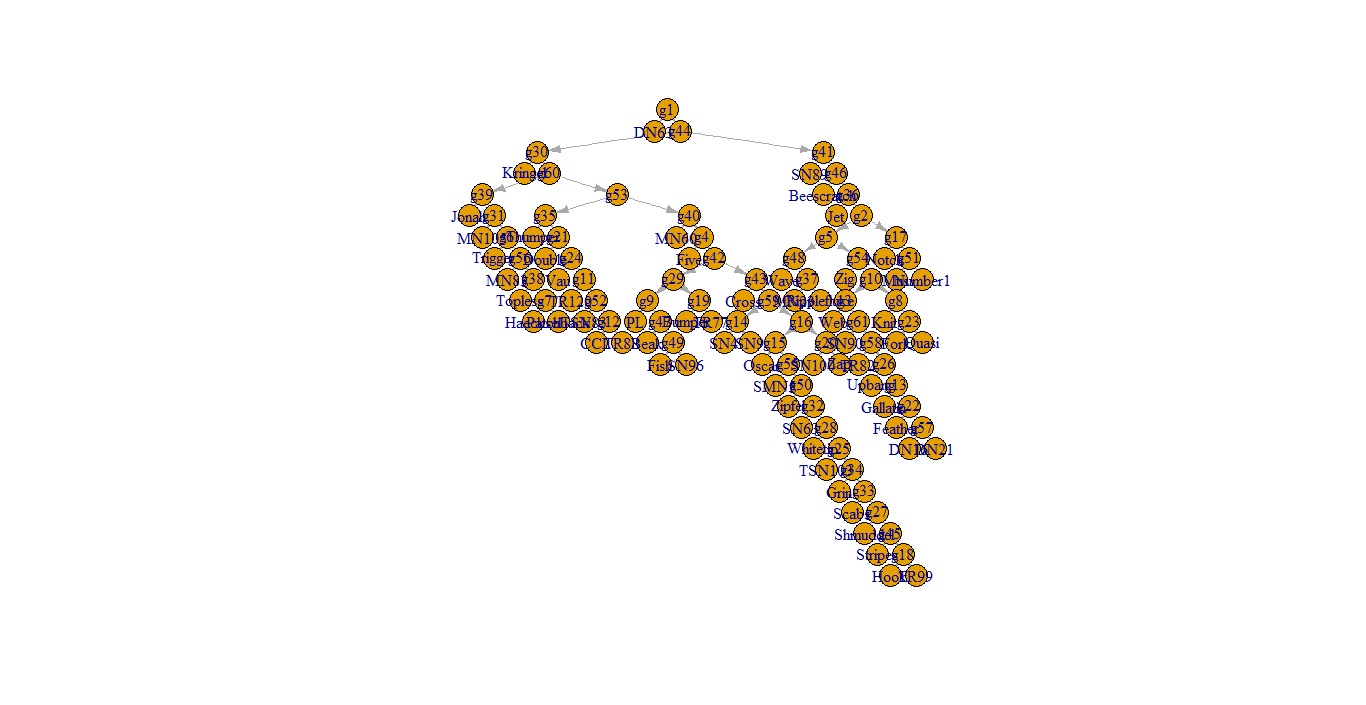
'- g46 p= 0.32 Beescratch

'- g36 p= 0.44 Zig Web SN90 TR82 Upbang Gallatin Feather

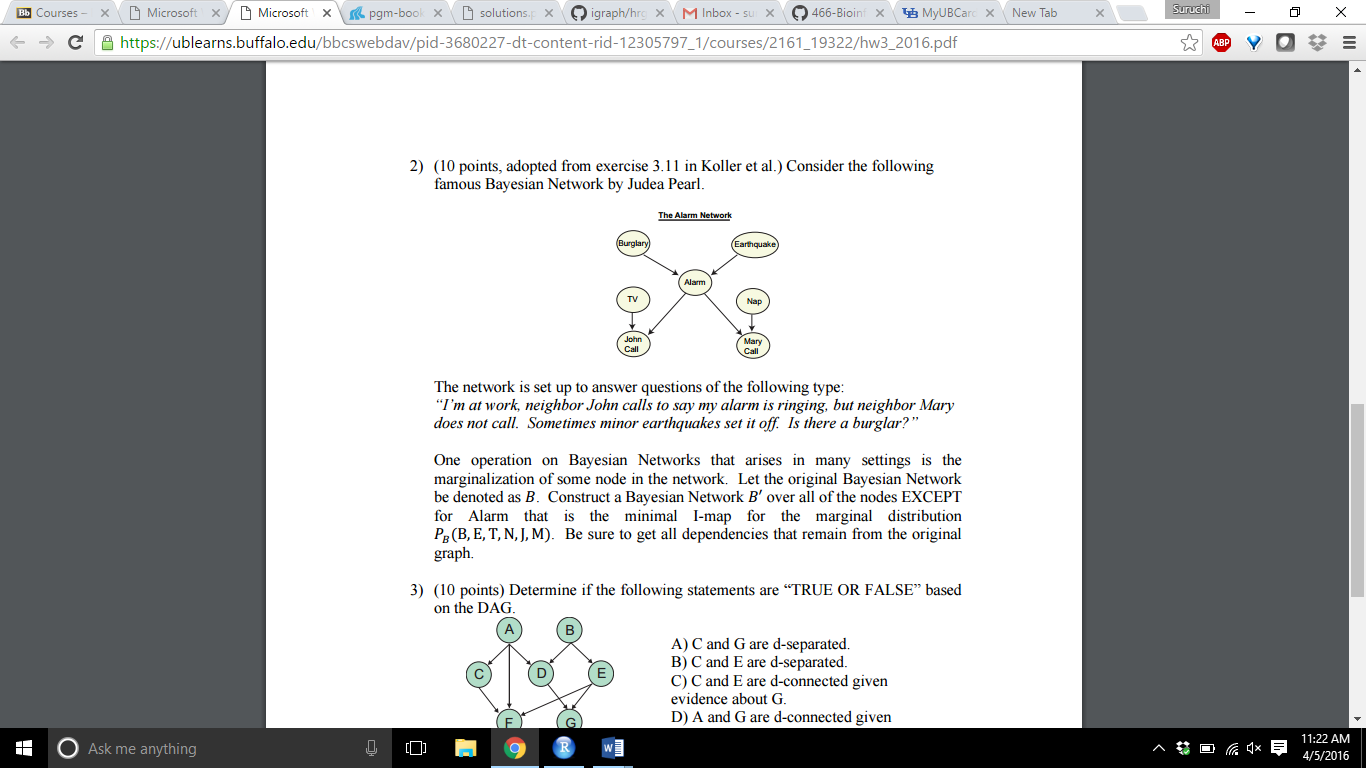
DN21 DN16 Knit Quasi Fork Wave Ripplefluke

MN23 Notch Number1 Mus Jet





Q2). The following famous Bayesian Network by Judea Pearl is given below -



Consider the Original Bayesian Network be B and the Bayesian Network to be constructed as B’–

Let Burglary be R, TV be T, John be J, Alarm be A, Earthquake be E, Nap be N and Mary be M.

Active trails in B are:

R → A → J, an edge has to be inserted between R and J, to assert an independence R Ʇ J.

J ← A → M, an edge has to be inserted between J and M.

T → J ← A → M, an edge has to be added between T and M, to assert an independence (T Ʇ |J, B, E)

An edge from each parent to each child of A, has to be added.

Moreover, the active trail in G between N and J, when M is observed, is still active in B’, due to the v-structure between N and J.

No other edges are required to be added.

Now to construct the Bayesian Network B’ – The minimal I-map construction algorithm is used.

The minimal I-map is a map without redundant edges.

The variables are ordered in the topological order in B.

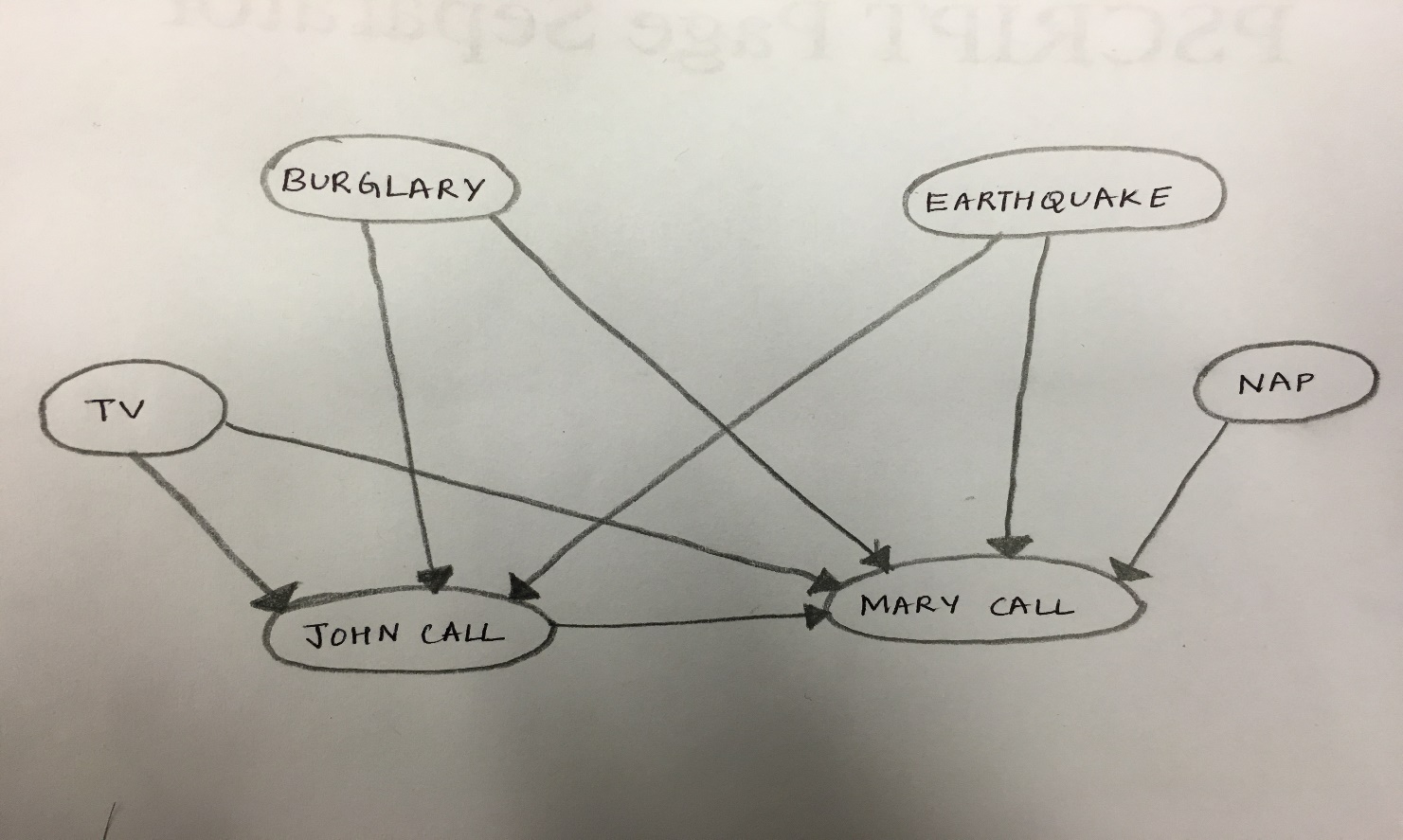
Without the loss of generality, the variables are renamed according to topological order:

. . , , ,. .,

Where is the alarm variable that is to be marginalized.

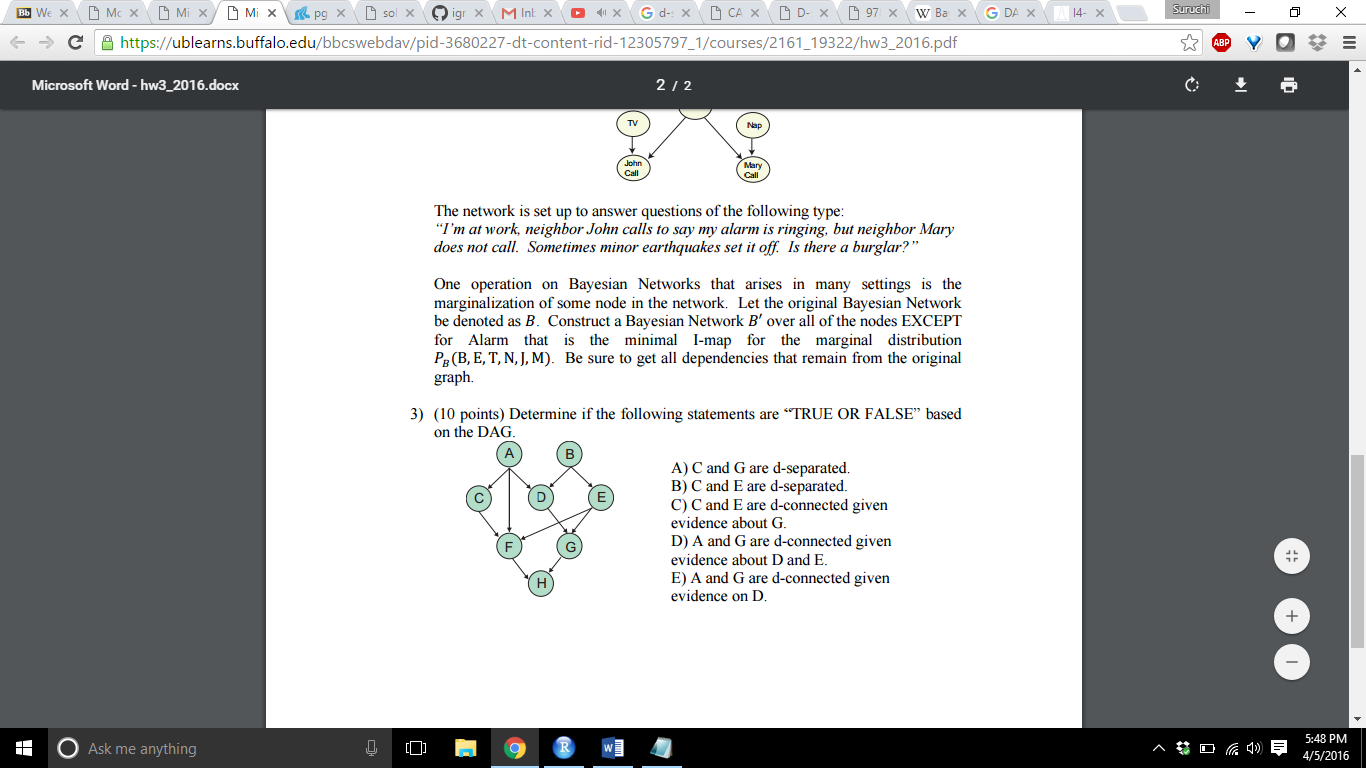
Add all the nodes to B’ using the above order, except the node and use the independencies to select a parent set for each variable. Add the variables . . , to the network B’, it results in the original network B. the parent set of variables from . . , will remain the same in constructed Bayesian Network B’, as it satisfies the local Markov assumption. The only variables that will be affected if the variable is marginalized (skipped) are those variables that has as a parent in B (‘s children). So for each child of , a new parent set is found. Now replace with a set of variables that act as a surrogate. Using the local Markov Assumption blocks the flow of information to the children variables, so the surrogate variables must block the same paths while being d-connected to the children variable of . Now consider the co-parents of , all the variables must appear in topological order. All the elements of the parent set are necessary to avoid creating independencies, so the siblings of the parents are to be added. These variables are required for the replacement of in parenting the variables of . These variables are a subset of Markov Blanket of , and they shield from other variables in the network, thus using them to replace . This ensures no other active trial can reach , and continue to the children variables.

Using this set of new parents, the local Markov Assumption is satisfied for the children variables in the Bayesian Network B’.

****

**Figure : Bayesian Network B’**

Q3) Consider the Directed Acyclic Graph shown below:



The d-separation is a criterion for deciding, whether a set X of variables is independent of another set Y, given a third set Z, from a graph.

The idea is to associate “dependence” with “connectedness” (i.e., the existence of a connecting path) and “independence” with “unconnectedness” or “separation.” To account for the orientations of the arrows the terms “d-separated” and “d-connected” (d connotes “directional”) are used.

To test whether 1 and 2 are d-separated by 3 in dag X, we need to consider every path between a node in 1 and a node in 2, and then ensure that the path is blocked by 3.

A path is blocked by 3 if at least one valve (node) on the path is ‘closed’ given 3. ν

A divergent valve or a sequential valve is closed if it is in 3, a convergent valve is closed if it is not on 3 nor any of its descendants are in 3.

a) C and G are d separated – FALSE

There is a serial connection A → D → G, allowing the transmission of information from A to G via D. So information can be transmitted from C to G via C → A → D → G, which shows that C and G are not d-separated. Hence the variables C and G are d-connected and not d-separated.

b) C and E are d separated – TRUE

The variables C and E are d-separated, since each path from C to E contains a converging connection, and since no evidence is available, each such connection will not allow transmission of information.

c) C and E are d connected given evidence about G – TRUE

The variables C and E contain a converging connection and there is no evidence available about other variables. But now since evidence is given about variable G, the variables C and E will become d-connected. Evidence about G, will show transmission of information from D to E via G, as a converging connection D → G ← E is present. Then the information from C to E can be transmitted via the diverging connection C ← A → D. So the information transmission occurs first via a divergent connection and then a convergent function,

C ← A → D → G ← E

d) A and G are d connected given evidence about D and E - FALSE

The variables A and G are d-separated, given evidence on D and E as it contains a convergent connection, D → G ← E and a straight path A →D → G, so no connection exists between A and G. Hence they are d-separated and not d-connected.

e) A and G are d connected given evidence on D – TRUE

The variables A an G are d- connected given evidence about variable D, as a straight connecting path between A and G exists, A → D → G.

------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------THE END-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------