SMU - Bayesian Networks Assignment
by José Ananías Hilario Reyes

i. Baseline Network

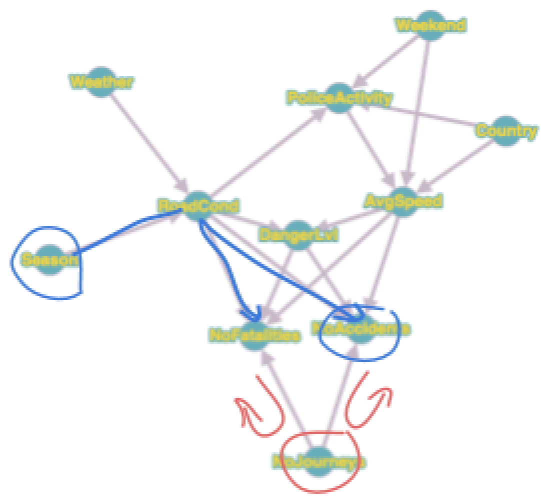
Baseline network was derived by intuition and by examination of the data, searching for plausible relations backed up by the evidence.

Independent nodes are: **NoJourneys**, **Weather**, **Season**, **Weekend**, **Country**.
Pr(**RoadCond** | **Season, Weather**)
Pr(**PoliceActivity** | **Weekend, Country, RoadCond**)
Pr(**AvgSpeed** | **Weekend, Country, PoliceActivity**)
Pr(**DangerLvl** | **RoadCond, AvgSpeed**)
Pr(**NoAccidents** | **RoadCond, DangerLvl, AvgSpeed, NoJourneys**)
Pr(**NoFatalities** | **RoadCond, DangerLvl, AvgSpeed, NoJourneys**)

ii. Information Flow Checks

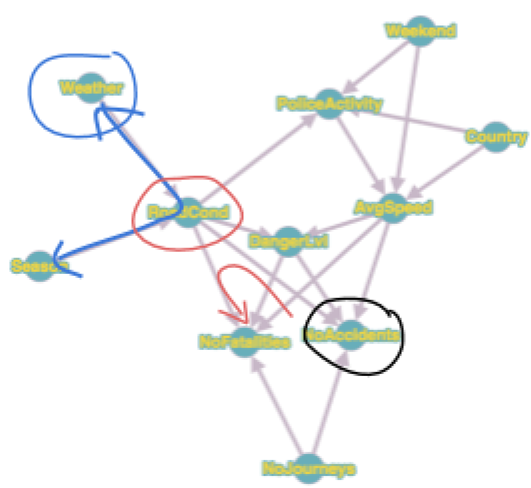
a) Season ⊥ NoFatalities | NoJourneys = False

There is dependence between Season and NoFatalities, given NoJourneys. In the model being used, NoJourneys is an independent variable, so its knowledge doesn't affect the relationships between the other nodes of the network. In natural state (nothing being observed), the network entails that NoFatalities is dependent on Season.



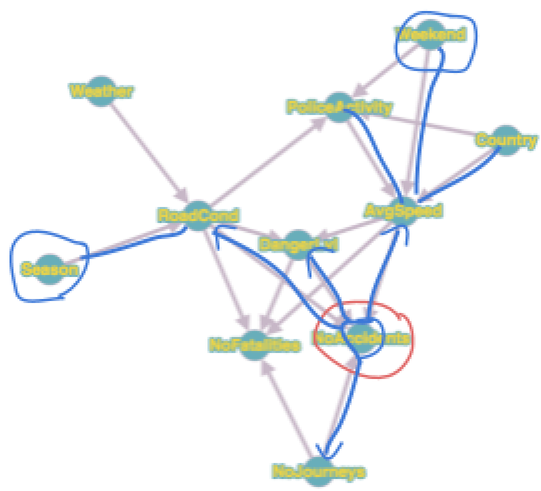
b) Weather ⊥ NoAccidents | RoadCond = True

Weather and NoAccidents are conditionally independent on RoadCond, since RoadCond itself is dependent on Weather. This means, when RoadCond is observed, no other information can be obtained from Weather, that is not already contained in this observation. Therefore, there is no information flowing in the graph from Weather to NoAccidents.



c) Season $\perp$ Weekend | NoAccidents = False

Season and Weekend are both independent variables in the network being considered. They are naturally independent. However, once NoAccidents is observed, the information trail flows between them. An intuitive example is that, if we know that the number of accidents is high, knowing the season which we are examining can give us information as to the day of the week it happened; ie, a high number of accidents in the summer, is more likely to happen on the holidays, for example.

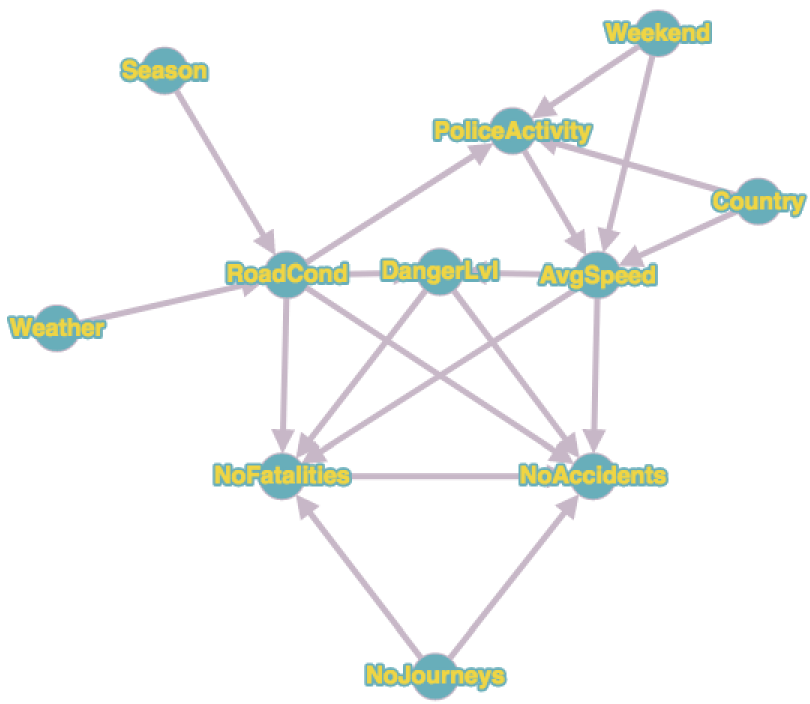


iii. CPT Interpretation

AvgSpeed	AvgSpeed(high)	AvgSpeed(high)	AvgSpeed(low)	AvgSpeed(low)
RoadCond	RoadCond(bad)	RoadCond(good)	RoadCond(bad)	RoadCond(good)
DangerLvl(high)	0.25443262411347517	0.04461200125667609	0.022363333517393705	0.013546202225447508
DangerLvl(low)	0.7455673758865248	0.9553879987433239	0.9776366648260629	0.9864537977745524

The interpretation of the CPT yielded can be seen quite forthrightly. Both an AvgSpeed which is high, and a bad RoadCond entail a higher danger level, each in its own measure. And rightly so, when RoadCond is good, and AvgSpeed is low we have the lowest probability of seeing a high DangerLvl; whereas when AvgSpeed is high, and RoadCond is bad, we have a dramatic change with respect to the rest, with a probability of 25% of having a high DangerLvl.

iv. Final Network



Final network was derived by exploring relations that were exposed by HillClimbing, search, and adding the relevant changes to the baseline network.

Only significant change added was a dependency of NoAccidents on NoFatalities.
Pr(**NoAccidents** | **RoadCond, DangerLvl, AvgSpeed, NoJourneys, NoFatalities**)

v. Jensen-Shannon divergence and total variation distance

Jensen-Shannon Divergence (JSD) = 0.26871878702172425
Total Variation Distance (TVD) = 0.011841848390446523