

## Assignment 2: Causal Inference

Giorgia Adorni (giorgia.adorni@usi.ch)

### 1 Structure of the Network

The problem modelled is a generic trip by car, influenced by factors such as the strike of public transport, road works or even the weather, and the delay that comes with it. The causal diagram that model this problem includes the following variables:

- **Weather:** weather during the journey that should be *sunny* or *rainy*.
- **Strike:** *true* if a strike of public transport takes place, *false* otherwise.
- **RushHour:** *true* if the time it's rush hour, *false* otherwise.
- **RoadConditions:** condition of the road floor, that depends on the weather, and should be *dry* or *wet*.
- **Humor:** humor of the driver, dependent on the weather, that is *good* or *bad*.
- **RoadWorks:** *true* if there are road maintenance works in progress, *false* otherwise. This variable depends on the conditions of the road.
- **Speed:** driving velocity, that should be *slow* or *fast*, dependent on the road condition, if it's rush hour and if there are road works.
- **Danger:** danger incurred during the trip, that should be *low* or *high*, dependent on the driving velocity, the road conditions and if it's rush hour.
- **Accident:** accident risk, that could be *low* or *high*, influenced by the danger, the humour of the driver and if there is a strike of the public transport.
- **Delay:** *true* if the trip is delayed, *false* otherwise.

The objective of the network is highlight hoe weather and humour impacts on travel safety. The graph could provide valuable indications about the correlation between the driver humour and a delayed trip, or for example between the weather and the risk of an accident and also on how the road conditions influences car crashes.

Each node is connected by an arrow to one or more other nodes upon which it has a causal influence. Most of the arcs orientation are self- explaining. An exception was made for the **Humor** variable, that influences the accident risk and is caused only by the weather and not for example by the delay.

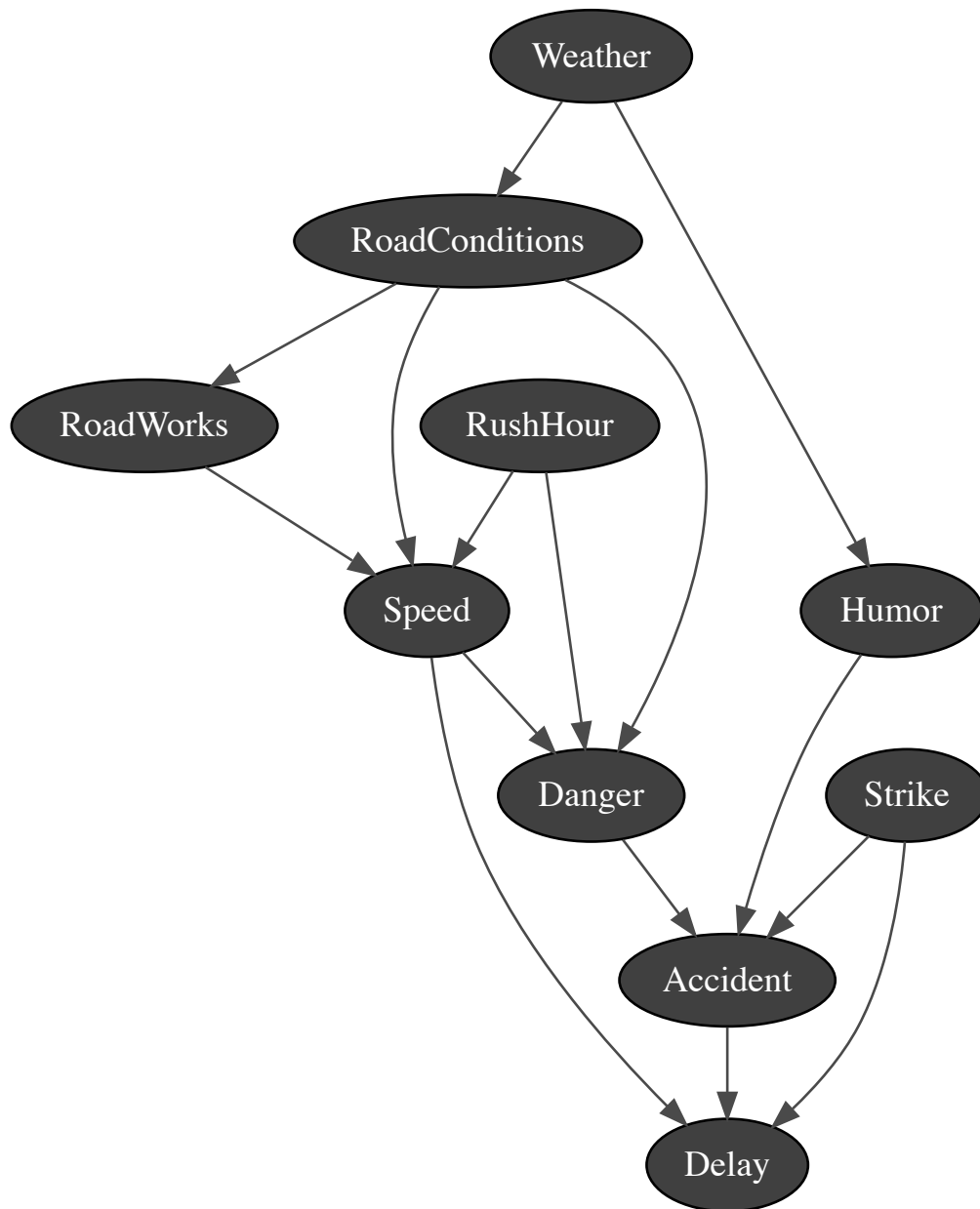


Figure 1: Bayesian Network

## D-Separation

Identify at least 4 couple of nodes (the node of each couple should be not directly linked to each other) and analyse their d-separation properties possibly conditioning on others.

Discuss how d-connected variables are in fact dependent in the real problem, while d-separated variables are instead independent in the real problem.

## 2 Conditional Probability Tables

Explain how do you fill the probability tables for the nodes. For instance: a. you have retrieved information from the internet (or other sources); b. you have estimated the CPTs from a database; c. you have relied on your personal experience/common sense.

Any causal model can be implemented as a Bayesian network. Bayesian networks can be used to provide the inverse probability of an event (given an outcome, what are the probabilities of a specific cause). This requires preparation of a conditional probability table, showing all possible inputs and outcomes with their associated probabilities.

For example, given a two variable model of Disease and Test (for the disease) the conditional probability table takes the form:

## 3 Causal Inference

Choose one pair of variables. The pair must be made up of a variable  $X$  with at least one parent and another variable  $Y$  of the graph such that there is (at least) a causal path from  $X$  to  $Y$ . For the pair  $(X,Y)$  perform: • Calculate the causal effect of  $X$  on  $Y$ . • Identify possible confounders between  $X$  and  $Y$ . • Would it be practically possible in your specific problem to perform also a randomized controlled study to disentangle the causal effect between the variables from their correlation? • Compute the ACE of  $X$  on  $Y$ . Choose another pair of variable  $(X,Y)$  (it can be also the previous one) and: • Choose another variable  $C$  such that it is possible to calculate the  $c$ -specific effect of  $X$  on  $Y$  and calculate it. • Identify a minimal set of variables that must be measured in order to estimate the  $c$ -specific effect of  $X$  on  $Y$ . • Choose a function  $g$  and compute the effect of the conditional intervention of  $X=g(C)$  on  $Y$ . Choose another pair of variable  $(X,Y)$  (it can be also the previous one) and: • Identify possible mediating variables between  $X$  and  $Y$  and calculate the CDE of  $Y$  changing the value of  $X$ .

## 4 Simulation

Suppose that you can't measure some parents of variable  $X$  chosen in every point of "Causal Inference". Repeat the "Causal Inference" part of the exercise considering this new situation.

## 5 Comment on the Results

What kind of experience have you got with this model? E.g., is the causal model responding in a sensible way to your queries? What should be changed/modified to make it more realistic?