CS440/ECE448 Lecture 26: Causal Networks

Mark Hasegawa-Johnson

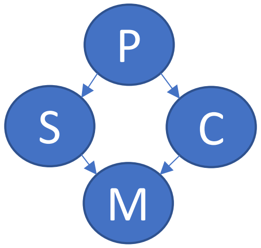
3/2022

Outline

- Review: Inference in Bayesian networks
- The do-calculus
- Some applications

Mail delivery in Whoville

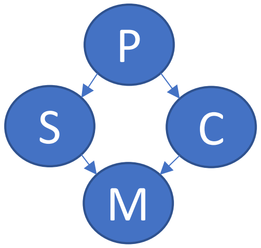
Rural_Free_Delivery_early_vehicle.JPG



- The village of Whoville sometimes pays its employees, and sometimes doesn't
- When the employees are paid, they are more likely to go out and shovel snow from the sidewalks
- When the employees are paid, they are more likely to charge the village's delivery vehicles
- When the delivery vehicle is charged, and the sidewalks are shoveled, the mailman is more likely to deliver mail

Mail delivery in Whoville

Rural_Free_Delivery_early_vehicle.JPG



Mary Lou Who loves to get mail, but she hates to go out to her mailbox and find it empty. Here are some questions she might ask:

- If the sidewalk has been shoveled, what's the probability that she has mail?
- If she goes out and shovels her sidewalk by herself, does that increase her chance of getting mail? By how much?

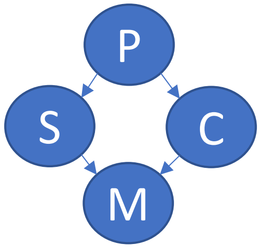

Snow shovelling.jpg

Review: Inference in Bayesian Networks

1. Find a path that includes all the variables of interest
2. Multiply conditional probabilities together to get the joint distribution of all variables on the path
3. Add the probabilities that correspond to different values of any variable you don't care about
4. Divide in order to get the conditional probability of the query given the observations

1. Find a path

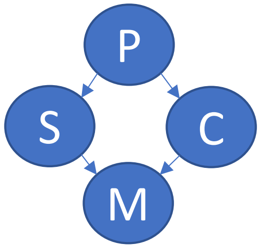
Rural_Free_Delivery_early_vehicle.JPG



- M is the query variable, S=T is the observation
- Their parents are P and C
- …in this case, the path we need to model is the whole graph, (P,S,C,M).

2. Multiply

Rural_Free_Delivery_early_vehicle.JPG

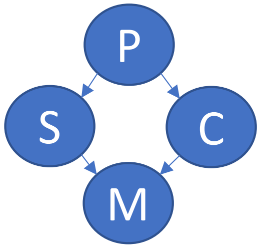


Suppose we have these probabilities:

- $P(P = T) = \frac{1}{2}$
- $P(S = T|P = T) = \frac{9}{10}, P(S = T|P = F) = \frac{1}{10}$
- $P(C = T|P = T) = \frac{9}{10}, P(C = T|P = F) = \frac{1}{10}$
- $P(M = T|S = T, C = F) = \frac{5}{10}, P(M = T|S = F, C = T) = \frac{5}{10},$
  $P(M = T|S = T, C = T) = \frac{9}{10}, P(M = T|S = F, C = F) = \frac{1}{10}$

2. Multiply

Rural_Free_Delivery_early_vehicle.JPG



We want the joint probability of all variables,

$$P(P,S,C,M) = P(P)P(S|P)P(C|P)P(M|S,C)$$

We observe that S=T, but we don't know the values of any of the other variables, so we need a table with 8 entries. For example,

$$P(P=T,S=T,C=T,M=T) = \left(\frac{1}{2}\right)\left(\frac{9}{10}\right)\left(\frac{9}{10}\right)\left(\frac{9}{10}\right) = \frac{9^3}{2000}$$

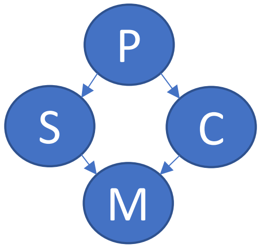
$$P(P=T,S=T,C=T,M=F) = \frac{9^2}{2000}$$

$$\vdots$$

$$P(P=F,S=T,C=F,M=F) = \frac{9\times 5}{2000}$$

3. Add

Rural_Free_Delivery_early_vehicle.JPG



We only really need the joint probability of S=T and M:

$$P(S=T,M)=\sum_{p=F}^{T}\sum_{c=F}^{T}P(P=p,S=T,C=c,M)$$

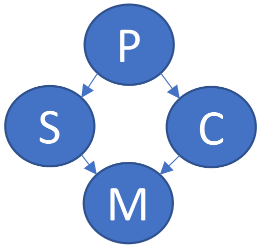
That gives a table with only two entries:

$$P(S=T,M=T)=P(T,T,T,T)+P(T,T,F,T)+P(F,T,T,T)+P(F,T,F,T)=\frac{9^3+9\times5+9+9\times5}{2000}$$

$$P(S=T,M=F)=P(T,T,T,F)+P(T,T,F,F)+P(F,T,T,F)+P(F,T,F,F)=\frac{9^2+9\times5+1+9\times5}{2000}$$

4. Divide

Rural_Free_Delivery_early_vehicle.JPG

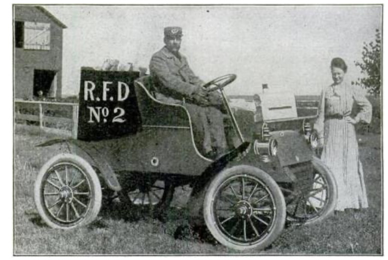


Finally, we just use the definition of conditional probability:

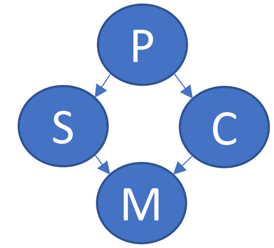
$$P(M = T | S = T) = \frac{P(M = T, S = T)}{P(M = T, S = T) + P(M = F, S = T)}$$

$$= \frac{9^3 + 9^2 + 9 + 9^2}{9^3 + 9\times5 + 9 + 9\times5 + 9^2 + 9\times5 + 1 + 9\times5} \approx 0.83$$

Mail delivery in Whoville



Rural_Free_Delivery_early_vehicle.JPG



Mary Lou Who loves to get mail, but she hates to go out to her mailbox and find it empty. Here are some questions she might ask:

- If the sidewalk has been shoveled, what's the probability that she has mail? ANSWER: 0.83
- If she goes out and shovels her sidewalk by herself, does that increase her chance of getting mail? By how much?

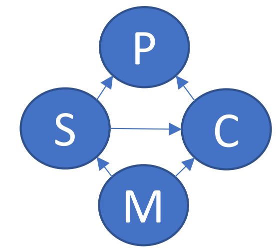
Snow shovelling.jpg

How can we answer the second question?

Outline

- Review: Inference in Bayesian networks
- The do-calculus
- Some applications

Non-Causal Bayesian Networks

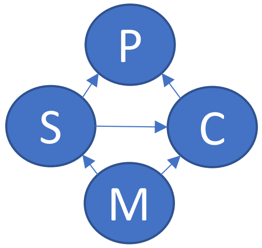


- It's possible to draw the arrows on a Bayesian network in non-causal order.
- Usually, non-causal ordering is less efficient. For example, S and C are not conditionally independent given knowledge of M (knowing that S=T makes it more likely that C=T), so we need to add an arrow between them.
- If we have all the necessary arrows, non-causal ordering should give the same answer to any inference problem. The multiply step is now

$$P(P,S,C,M) = P(M)P(S|M)P(C|M,S)P(P|S,C)$$

…and then the add and divide steps work the same as before.

Non-Causal Bayesian Networks

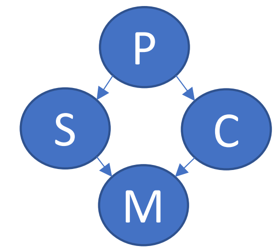


Using a non-causal network, however, we can't answer questions like:

"If I act on the system by changing the value of one variable, what will happen to the other variables?"

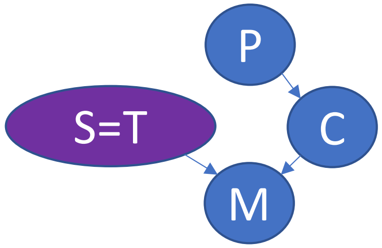
We lose the ability to represent causal changes. For example, if Mary Lou puts a letter in her own mailbox, that sets M=T, but it does not cause her sidewalk to get shoveled.

Causal Networks



- A **causal network** is a Bayesian network in which all arrows point from a cause to an effect.
- The benefit of a causal network is that we can ask causal questions.
- For example: If Mary Lou goes out and shovels her sidewalk by herself, does that increase her chance of getting mail? By how much?
- We can answer that question now, because our model includes the parameters $P(M = T|S = T, C = T)$ (which answers the question if $C = T$) and $P(M = T|S = T, C = F)$ (which is the answer if $C = F$).

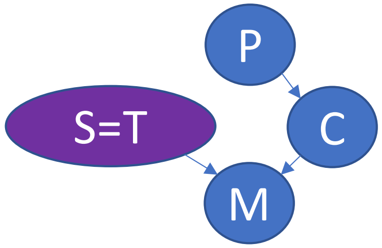
Do-calculus



In order to analyze the effect of some action (say, setting $S = T$), network theorists use the "do" operation (for example, $\mathrm{do}(S = T)$). The do operation includes the following steps:

1. Separate the action variable from its parents.
2. Assign it the specified value (e.g., T)

Do-calculus



Now we have

$$P(P=T)=\frac{1}{2}$$

$$P(C=T)=\sum_{p=F}^{T}P(P=p)P(C=T|P=p)=\left(\frac{1}{2}\right)\left(\frac{9}{10}\right)+\left(\frac{1}{2}\right)\left(\frac{1}{10}\right)=\frac{1}{2}$$

$$P(M=T|\mathrm{do}(S=T))=\sum_{c=F}^{T}P(C=c)P(M=T|S=T,C=c)$$

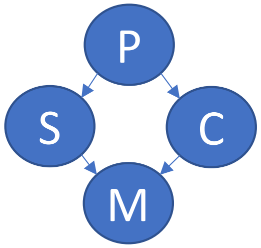
$$=\left(\frac{1}{2}\right)\left(\frac{9}{10}\right)+\left(\frac{1}{2}\right)\left(\frac{5}{10}\right)=\frac{7}{10}$$

Mail delivery in Whoville

Rural_Free_Delivery_early_vehicle.JPG

- If the sidewalk has been shoveled, what's the probability that Mary Lou has mail?

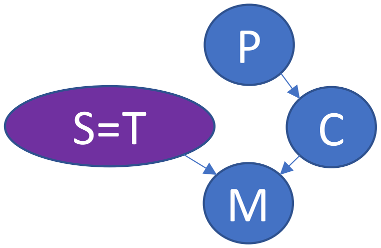
  ANSWER: $P(M = T|S = T) = 0.83$



Snow shovelling.jpg

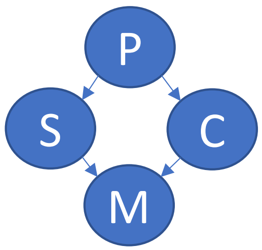
- If she goes out and shovels her sidewalk by herself, what is the probability that Mary Lou gets mail?

  ANSWER: $P(M = T|\text{do}(S = T)) = 0.7$

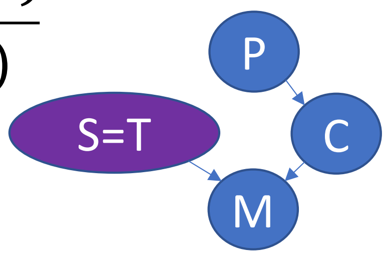


Why are they different?

$$P(M = T|S = T) = \frac{\sum_{p,c} P(P)P(S = T|P)P(C|P)P(M = T|S = T, C)}{\sum_{p,c,m} P(P)P(S = T|P)P(C|P)P(M|S = T, C)}$$

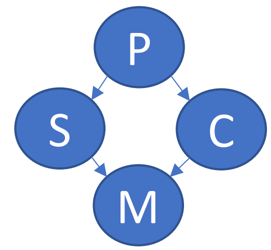


$$P(M = T|\mathrm{do}(S = T)) = \frac{\sum_{p,c} P(P)P(C|P)P(M = T|S = T, C)}{\sum_{p,c,m} P(P)P(C|P)P(M|S = T, C)}$$

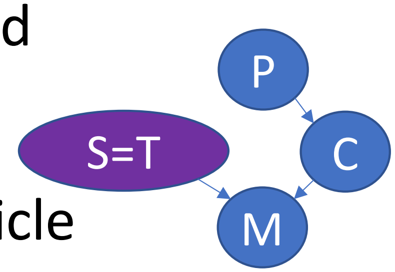


Why are they different?

$$P(M=T|S=T)=\frac{\sum_{p,c}P(P)P(S=T|P)P(C|P)P(M=T|S=T,C)}{\sum_{p,c,m}P(P)P(S=T|P)P(C|P)P(M|S=T,C)}$$

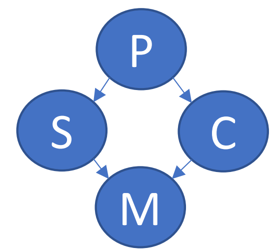


- $P(S=T|P=T)=\frac{9}{10}$, but $P(S=T|P=F)=\frac{1}{10}$
- The shoveled snow is a kind of signal: observing the shoveled snow, we conclude that the employees were probably paid today.
- Knowing that the employees were paid implies that the vehicle was probably also charged, which makes it more likely that Mary Lou has mail.

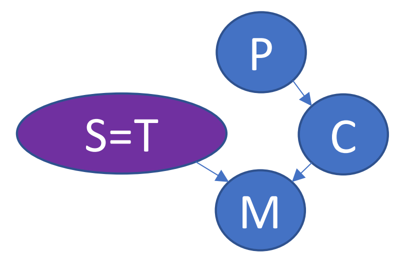


Why are they different?

$$P(M=T|S=T)=\frac{\sum_{p,c}P(P)P(S=T|P)P(C|P)P(M=T|S=T,C)}{\sum_{p,c,m}P(P)P(S=T|P)P(C|P)P(M|S=T,C)}$$



$$P(M=T|\mathrm{do}(S=T))=\frac{\sum_{p,c}P(P)P(C|P)P(M=T|S=T,C)}{\sum_{p,c,m}P(P)P(C|P)P(M|S=T,C)}$$



- The shoveled snow is no longer a signal, because Mary Lou did it herself.
- The vehicle is equally likely to have been charged or not charged.
- The mail is only 70% likely to arrive.

Mail delivery in Whoville

Rural_Free_Delivery_early_vehicle.JPG

Snow shovelling.jpg

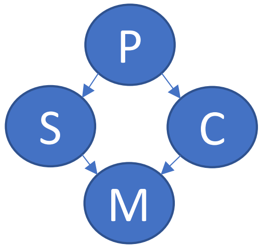
Wait… did shoveling the snow hurt Mary Lou's chances?

No. It just took away the value of the shoveled-snow signal.

- If the sidewalk has been shoveled, what's the probability that Mary Lou has mail?

ANSWER: shoveled snow is a signal.

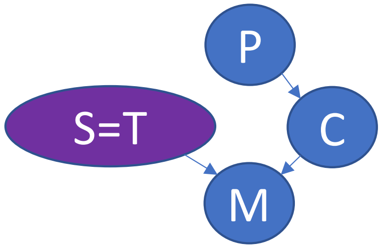
$$P(M = T|S = T) = 0.83$$



- If she goes out and shovels her sidewalk by herself, what is the probability that Mary Lou gets mail?

ANSWER: shoveled snow is not a signal.

$$P(M = T|\mathrm{do}(S = T)) = 0.7$$



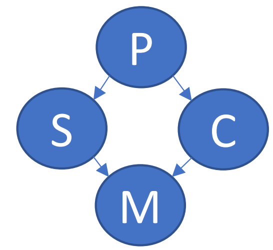
Mail delivery in Whoville

Rural_Free_Delivery_early_vehicle.JPG

Snow shovelling.jpg

Instead, we should compare these two cases:

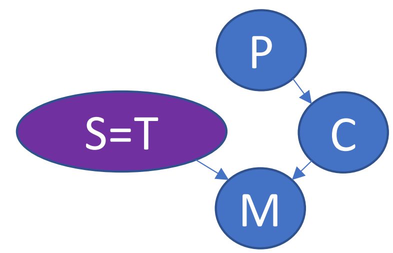
- If Mary Lou doesn't shovel her own snow, what's the probability she gets mail?



ANSWER:

$$P(M=T) = \frac{\sum_{p,s,c} P(P)P(S|P)P(C|P)P(M=T|S,C)}{\sum_{p,s,c,m} P(P)P(S|P)P(C|P)P(M|S,C)} = 0.5$$

- If Mary Lou shovels her own snow, what's the probability she gets mail?



ANSWER:

$$P(M=T|\text{do}(S=T)) = 0.7$$

Outline

- Review: Inference in Bayesian networks
- The do-calculus
- Some applications

Choosing which ads to show to a user

(©Bottou et al., Counterfactual Reasoning and Learning Systems, ICML 2013)

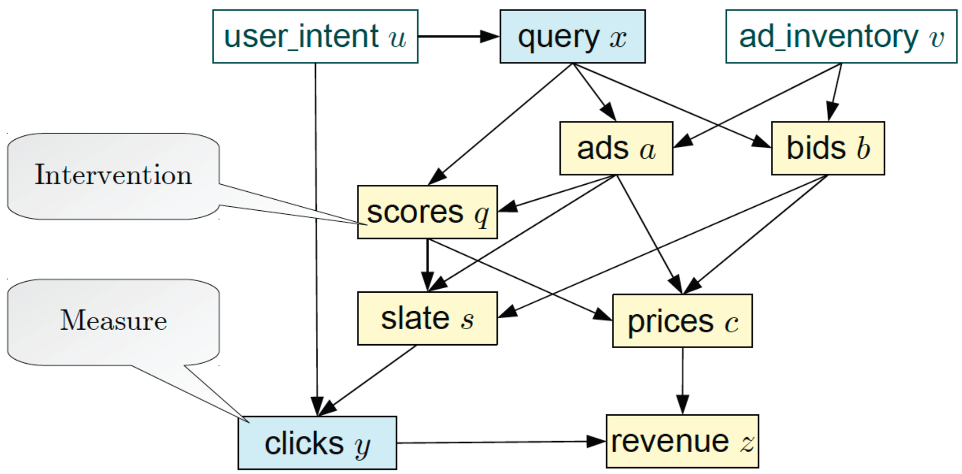
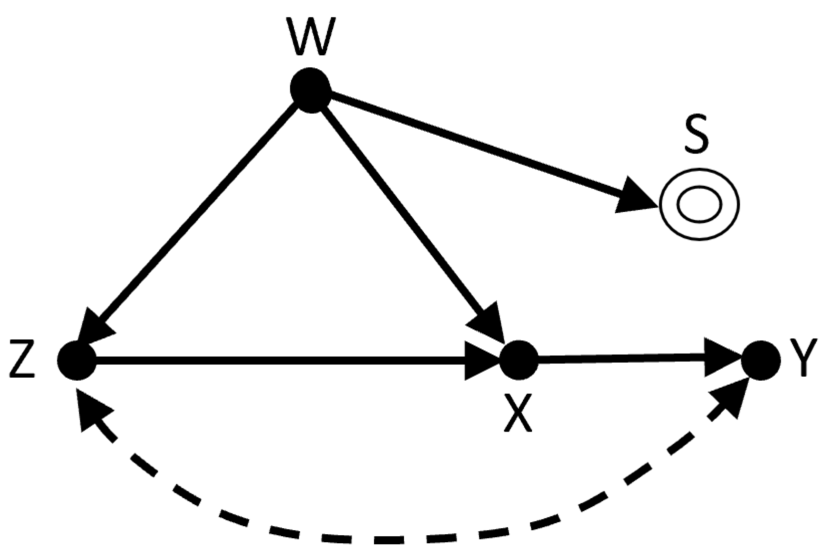


Figure 12: Estimating which average number of clicks per page would have been observed if we had used a different scoring model.

Figuring out whether students would score well if they all studied equally hard

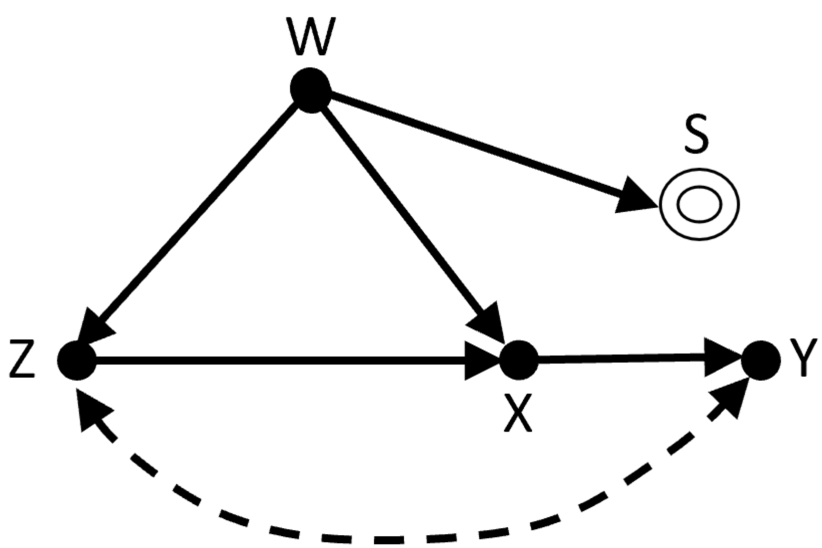
(© Ben Ogorek, https://towardsdatascience.com/do-calculus-and-continuous-distributions-866fb6a963bc)



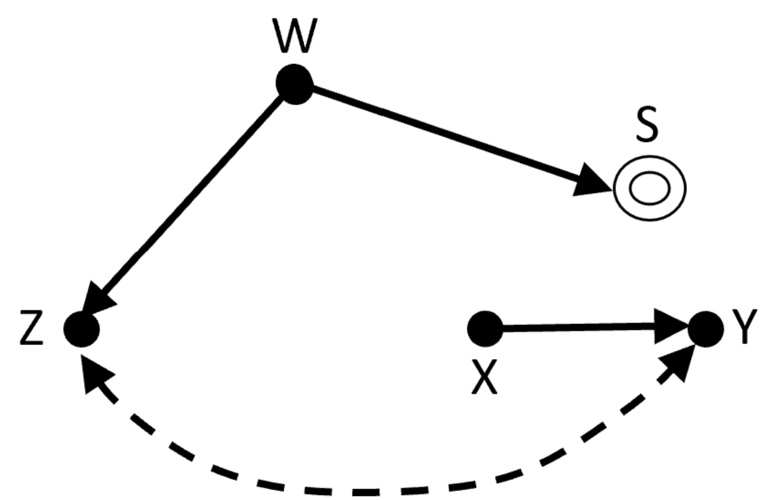
- S = is student enrolled in the course?
- W = # hours a student has available until the test
- Z = how much student enjoys the material
- X = # hours student spends studying for the test
- Y = grade received

Figuring out whether students would score well if they all studied equally hard

(© Ben Ogorek, https://towardsdatascience.com/do-calculus-and-continuous-distributions-866fb6a963bc)



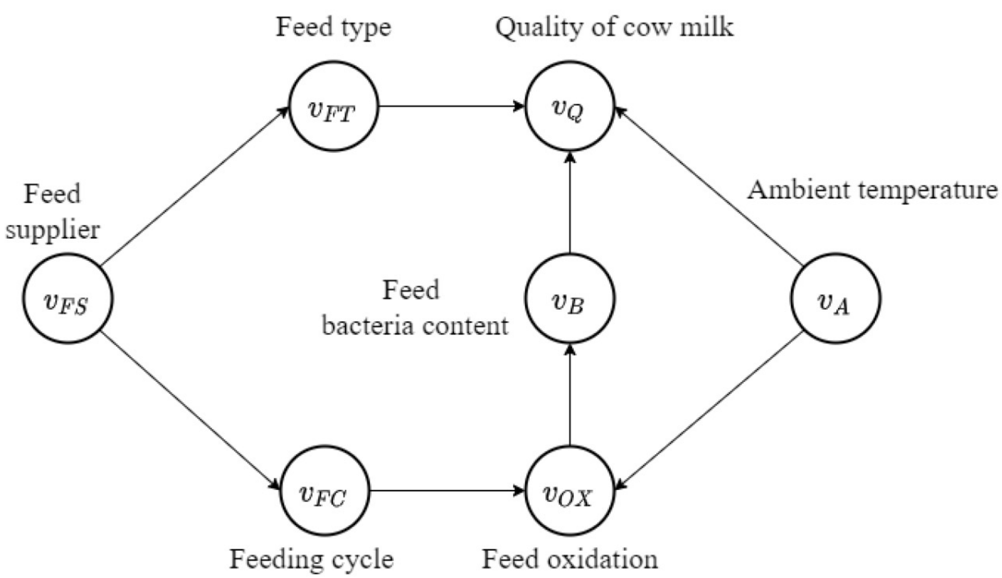
Factual graph: students who enjoy the class more tend to spend more time studying, hence get better grades.



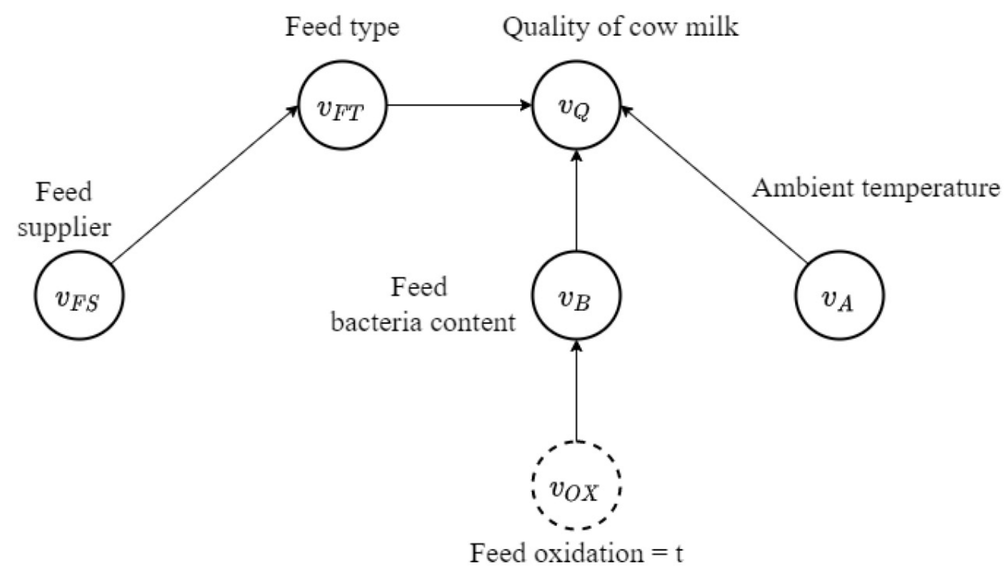
Counterfactual graph: what if all students spent the same amount of time studying?

Application of do-calculus for the β-robust scheduling method in a one-machine job environment

(© Quinten Rademakers, https://repository.tudelft.nl/islandora/object/uuid:2999f6d0-ebd0-4d1d-9651-583433f8b31f)



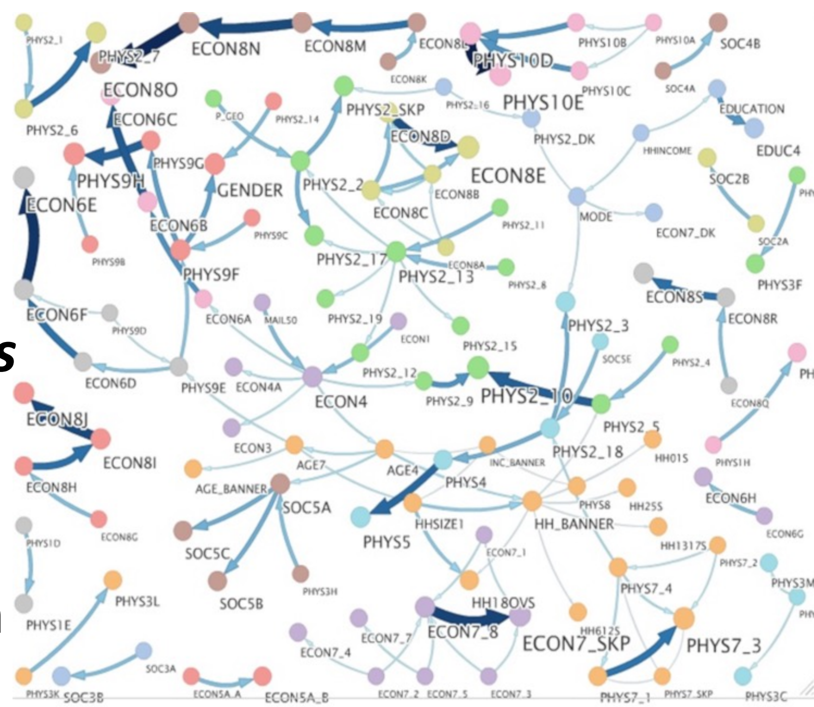
Model of a dairy farm



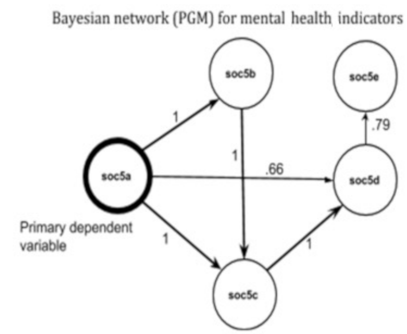
What would happen if we left the air seal on the feed container open all the time?

Learning the Mental Health Impact of COVID-19 in the United States With Explainable Artificial Intelligence: Observational Study
 https://mental.jmir.org/2021/4/e25097

"The mental health consequences of the COVID-19 pandemic have been substantial… This study addresses this gap through the use of ***Bayesian networks (BNs)***, an ***explainable artificial intelligence*** approach that captures the joint multivariate distribution underlying large survey data"



(a)



(b)

Outline

- Review: Inference in Bayesian networks

$$P(M=T|S=T)=\frac{\sum_{p,c}P(P)P(S=T|P)P(C|P)P(M=T|S=T,C)}{\sum_{p,c,m}P(P)P(S=T|P)P(C|P)P(M|S=T,C)}$$

- The do-calculus

$$P(M=T|\mathrm{do}(S=T))=\frac{\sum_{p,c}P(P)P(C|P)P(M=T|S=T,C)}{\sum_{p,c,m}P(P)P(C|P)P(M|S=T,C)}$$

- Some applications: Interventions, Explainable AI