Causal Inference: Algorithms

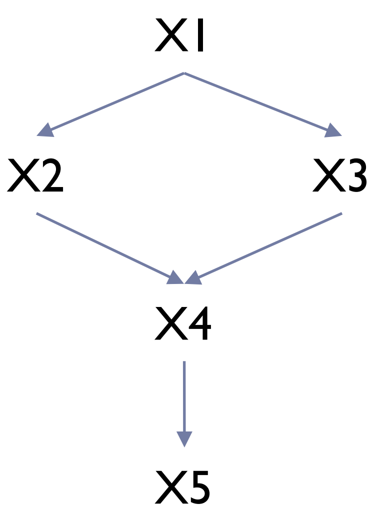
Michał Kot

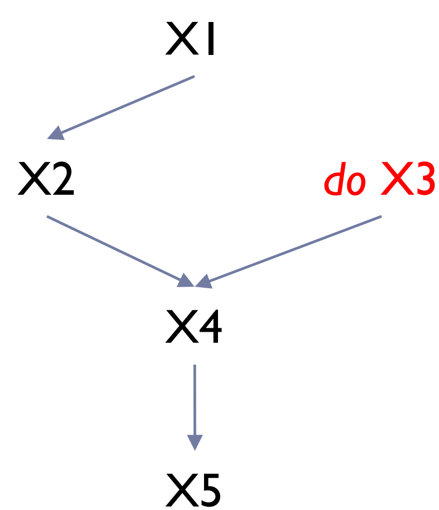
Goal (and agenda)

- Intuition
- Algorithms for causal inference in observational data
  - No latent variables
  - With latent variables
- Examples in R

Quick reminder

- Data
  - Observational vs interventional
- Inference
  - Observational vs interventional vs counterfactual
- Graphs representing mechanisms, independencies, interventions e.g:





Counterfactuals

- Three levels of inference:
  - observational
    - „patient takes drug A, will he recover?"
  - interventional
    - „what will happen if I will prescribe patient drug A?"
  - counterfactual
    - „patient took drug A and recovered, what would happen if took drug B?"
    - Recovery resulting from A / independently of A / despite A
- Are counterfactuals necessary?
  - Repeatable decisions
  - What would happen without treatment? Question linked to another one regarding % of patients who will die when treated

Counterfactual inference dilemma. Neyman-Rubin model

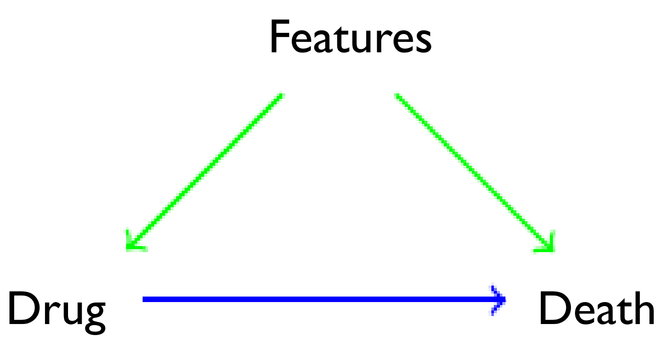
- Patients have been randomly assigned to treatment (X=1) or not (X=0)
- Recovery has been observed (Y=1) or not (Y=0)

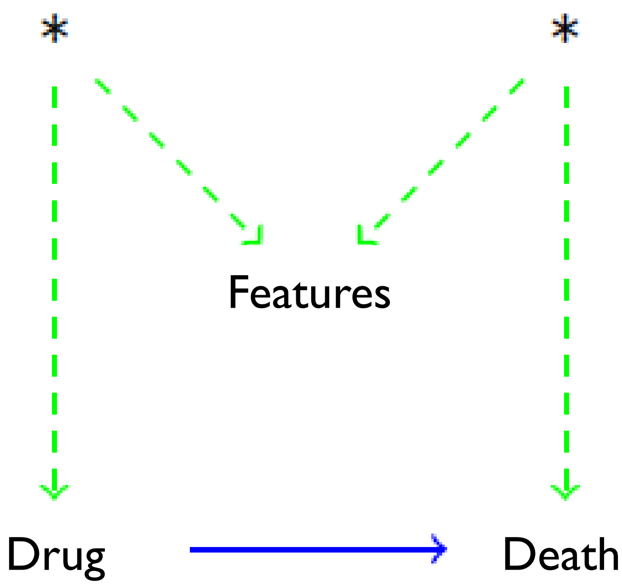
	Y=0	Y=1
X=0	15%	10%
X=1	30%	45%

- For patient X=0, Y=1, what would happen if X=1?

Causal inference – why we use it?

- Knowledge
- Information regarding intervention impact on other variables…
- … and its strenght based on proper *deconfounding*





Methods of causal inference

- Correlation and time precedence (Granger causality, *predictive causality*), but
  - Barometer and storm
  - Rational expectations

- Interventions e.g randomization in treatment (Neyman-Rubin model), but
  - Potentially not ethical randomization
  - Experiments non applicable to economics (most often)
  - Observational data may be useless here – but is often the only available source of information

- Therefore, is it possible to detect causal relationships in observational data (not a time-series)?
  - Observational data – source of information regarding conditional independencies
    - In the next few slides we forget about inference based on random samples (very important task in practical usage)

An example #1

A zoologist travels to Africa to study the natural breeding environment of giraffes. While there, he notices a type of tall tree that produces a special fruit that only grows at the top of the tree. He also notices that giraffes that frequently eat this fruit appear to be stronger and taller than those who cannot reach the fruit. He concludes that the fruit contains rich nutrients which make the giraffes that eat the fruit grow stronger and taller.

An example #1

A zoologist travels to Africa to study the natural breeding environment of giraffes. While there, he notices a type of tall tree that produces a special fruit that only grows at the top of the tree. He also notices that giraffes that frequently eat this fruit appear to be stronger and taller than those who cannot reach the fruit. He concludes that the fruit contains rich nutrients which make the giraffes that eat the fruit grow stronger and taller.

1. When a giraffe frequently eats this special fruit, it grows stronger and taller.
2. The nutrients in the fruit can help the giraffe grow stronger and taller.
3. Both A and B are correct.
4. The result is not sufficient to demonstrate that eating the fruit causes a giraffe to grow stronger and taller.
5. None of the above statements is reasonable.

Common fallacies

A is correlated to B, therefore A causes B

Possible answers

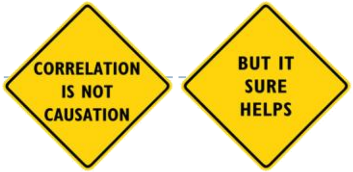
1. A may be causing B
2. B may be causing A
3. Unknown factor C may be causing both A and B
4. Possible combination of above, e.g bidirectional relationship
    1. E.g Temperature vs. Pressure in Ideal Gas Law when volume in constant
5. Coincidence

An example #2

- How do we know it is not firemen appearance that cause fire?

- Because they do not deal only with fire?
  - Yet sometimes smoker does not have lung cancer
- Because I may make him appear and there will be no fire!
  - Yet it is interventional
- Because firemen appearance is also correlated with other event (e.g flood) independent of fire
  - It requires assumption that firemen appearance does not cause both fire and flood and that fire and flood are independent

- Therefore
  - Causality detection requires analysis of variable triples
  - We assume: ***no causation without correlation***



Intuition (1)

- Basic concept: A is a cause of B, if had A not happened B would not have happened (Lewis 1973)
- But consider an example:
  - Suzy and Billy pick up rocks and throw them at a bottle. Suzy's rock gets there first, shattering the bottle. Since both throws are perfectly accurate, Billy's would have shattered the bottle had it not been preempted by Suzy's throw
- Therefore, is Suzy's throw a cause of bottle shattering or not?
- It is so-called but-for condition

Intuition (2) – stability

- We want three variables A, B, C, so that
  - A and B are dependent
  - B and C are dependent
  - A and C are independent

- Natural example A → B ← C
  - So that independent causes of common result
  - Yet consider Suzy and Billy example. Are both of A and C sufficient for B to happen?

- Possible example, that requires special parameter assignment
  - B∈{♣, ♦, ♥, ♠}
  - A – colors (♣, ♦)
  - C – black color
  - For A ← B → C and A $\perp\!\!\!\perp$ C to happen we require parameter assignment

Intuition (3) – minimality

- Think of an example: bottle has shattered because Suzy has thrown a rock and sneezed
- Was sneezing a cause of bottle shattering?
- Therefore we should prune inessential elements of a cause
- We want our causal relationships to be minimal, only relevant elements constitute the cause

Types of causes

- Necessary causes:
  - If X is a necessary cause of Y, then the presence of Y necessarily implies the presence of X with a probability of 100%. The presence of X, however, does not imply that Y will occur.
- Sufficient causes:
  - If X is a sufficient cause of Y, then the presence of X necessarily implies the presence of Y with the probability of 100%. However, another cause Z may alternatively cause Y. Thus the presence of Y does not imply the presence of X.
  - For example, if it is sunny outside, then it is daytime. It being sunny is a sufficient cause for one to conclude that it is daytime. But just because it is daytime does not necessarily mean it is sunny outside.
- Contributory causes:
  - If X is a contributory cause of Y, it means the presence of X makes possible the presence of Y, but not with the probability of 100%. In other words, a contributory cause may be neither necessary nor sufficient but it must be contributory.
  - For example, stubbing my toe causes pain. Stubbing my toe is a contributory cause to being in pain because I could be in pain from a headache or sore throat instead.

Halpern-Pearl definition of causality - foundations

- We have a causal model M, that is a pair (S, F).
  - S is a tuple - a list of elements U,V and R:
    - U - exogenious variables, that we set to u, so that U=u
    - V – endogenious variables
    - R – possible values of all variables in U and V (e.g if Y is binary then R(Y) = {0,1}
  - F is a function defining each endogenius variable X in V
    - F determines value of X, given all other variables
- We restrict M to be acyclic:
  - If A is a cause of B, then B is not a cause of A
- We may think of M as of acyclic directed graph (DAG)
  - If A is a cause of B, then A is a parent node of B

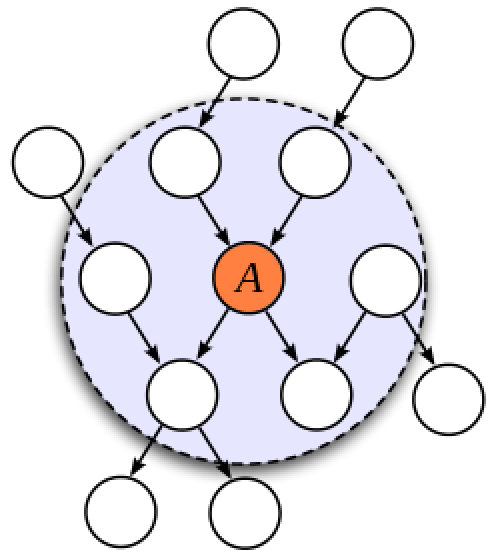
Halpern-Pearl definition of causality – definition itself

- X is an actual cause of Y if:
  - X and Y actually happened (AC1)
  - X is a minimal cause of Y (AC3)
  - Two more requirements:
    - Extended but-for condition: there exisits a subset of endogenious variables and possible setting of its values, so that given X not happened Y would not have happened. In our case if we set Billy to not throw, then if Suzy had not thrown a rock, bottle would not have shattered (AC2a)
    - Sufficiency requirement: if we set all variables to values that we actually observed, Y would happened (AC2b)
- Detailed discussion in Halpern, Pearl „A modification of Halpern-Pearl definition of causality", 2015

How to create DAG if we have observational data?

- Goal(s):
  - to build a minimal, stable undirected acyclic graph presenting dependencies between features based on data
  - to build a structural equation model to detect direction and strength of a relationship based on estimated DAG
- Interesting facts:
  - If A is a cause of B, then A is a parent node of B in DAG
  - B is independent of all its ascendants in DAG given its parent nodes
  - B is independent of any variables that are not its Markov blanket

What is a Markov blanket in DAG?



Markov blanket of A is:

- Its parents
- Its children
- Its children other parents

Directed Acyclic Graph build algorithm (PC, Peter Sprites and Clark Glymour)

- Assumptions:
  - No cycles and latent variables
  - Stability – We have a single skeleton
  - Minimality – we prune all not neccessary connections
- We analyze conditional dependencies to obtain
  - Graph skeleton
  - Edges orientation
- Algorithm:
  - Start with complete graph, where all variables are connected
  - For all pairs of variables A and B, try to find a set of variables S (it may be empty), so that $A \perp\!\!\!\perp B \mid S$
    - If S does not exist, A and B should remain connected
  - If A and B are not connected and have a common neighbour C, check if $C \in S_{AB}$
    - If not, then A→C←B
  - Orient as many edges as possible, without creating cycles or structure (…→…←…)

Limitations in detection of causality

- Based on observational data, some graphs may be equivalent, e.g:

A → B → C vs

A ← B ← C vs

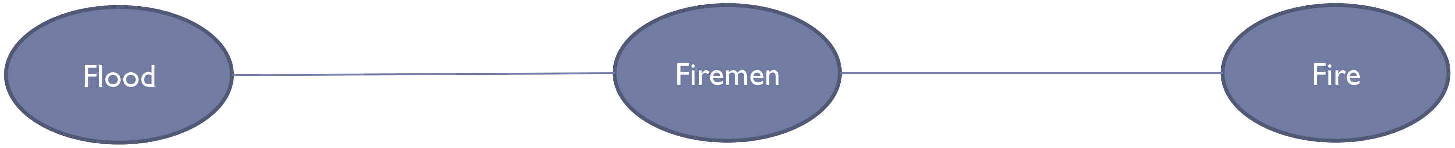
A ← B → C

A → B vs

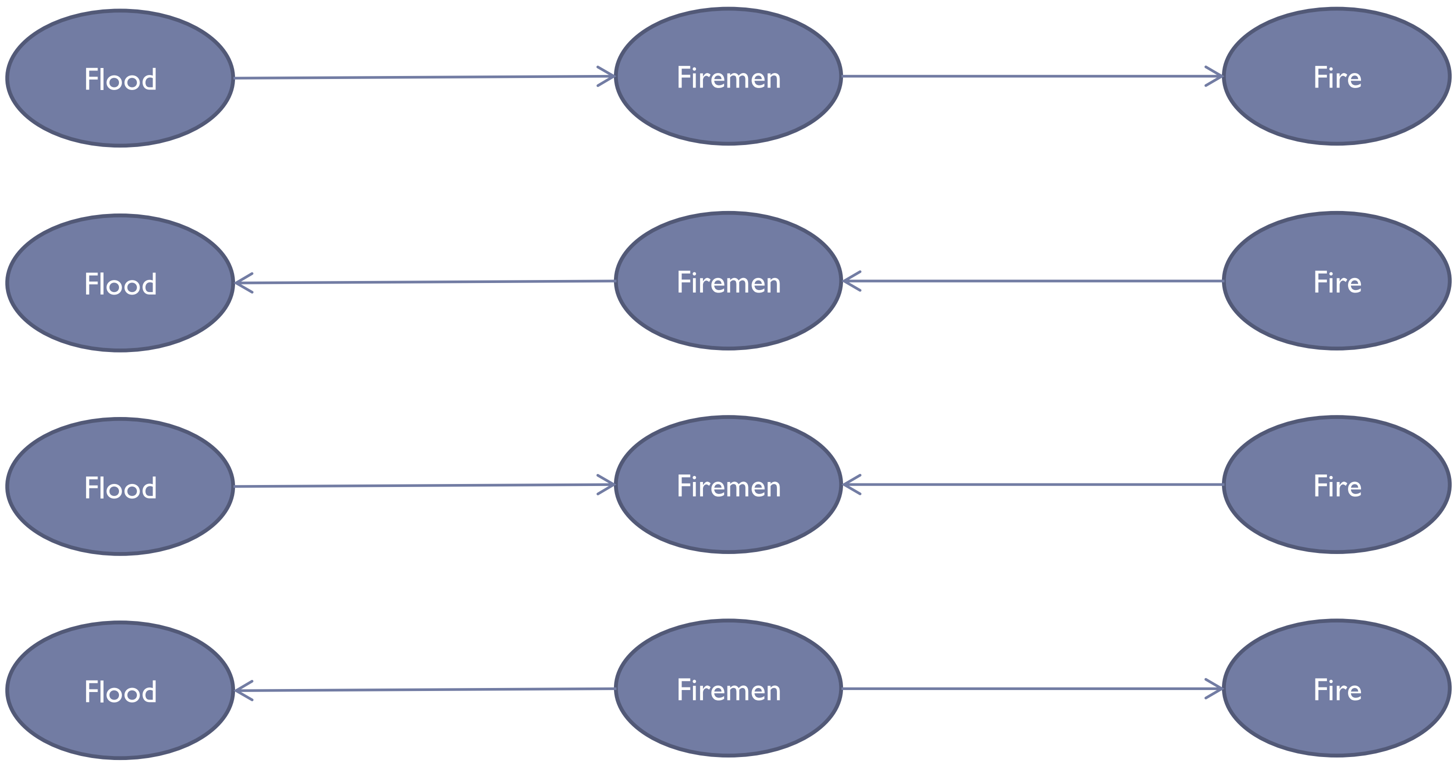
A ← * → B

PC for firemen, fire and flood

- Observations:
  - flood $\perp\!\!\!\perp$ fire
  - flood $\not\perp\!\!\!\perp$ firemen
  - fire $\not\perp\!\!\!\perp$ firemen

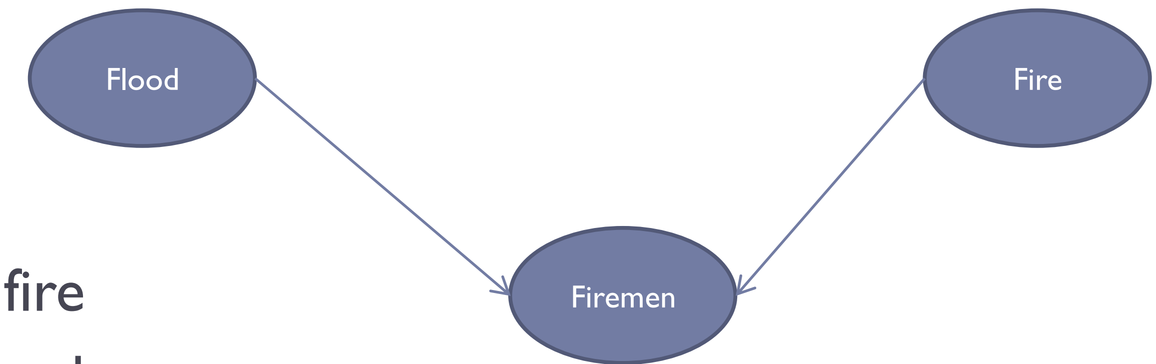


Possible v-structures



PC for firemen, fire and flood cont.

- **Observations:**
  - flood $\perp\!\!\!\perp$ fire
  - flood $\not\perp\!\!\!\perp$ firemen
  - fire $\not\perp\!\!\!\perp$ firemen
  - flood $\not\perp\!\!\!\perp$ firemen $|$ fire
  - fire $\not\perp\!\!\!\perp$ firemen $|$ flood
  - flood $\not\perp\!\!\!\perp$ fire $|$ firemen



- **Algorithm**
  - Connects **flood** and **firemen**
  - Connects **fire** and **firemen**
  - Orients edges towards **firemen**

With latent variables (algorithm FCI)

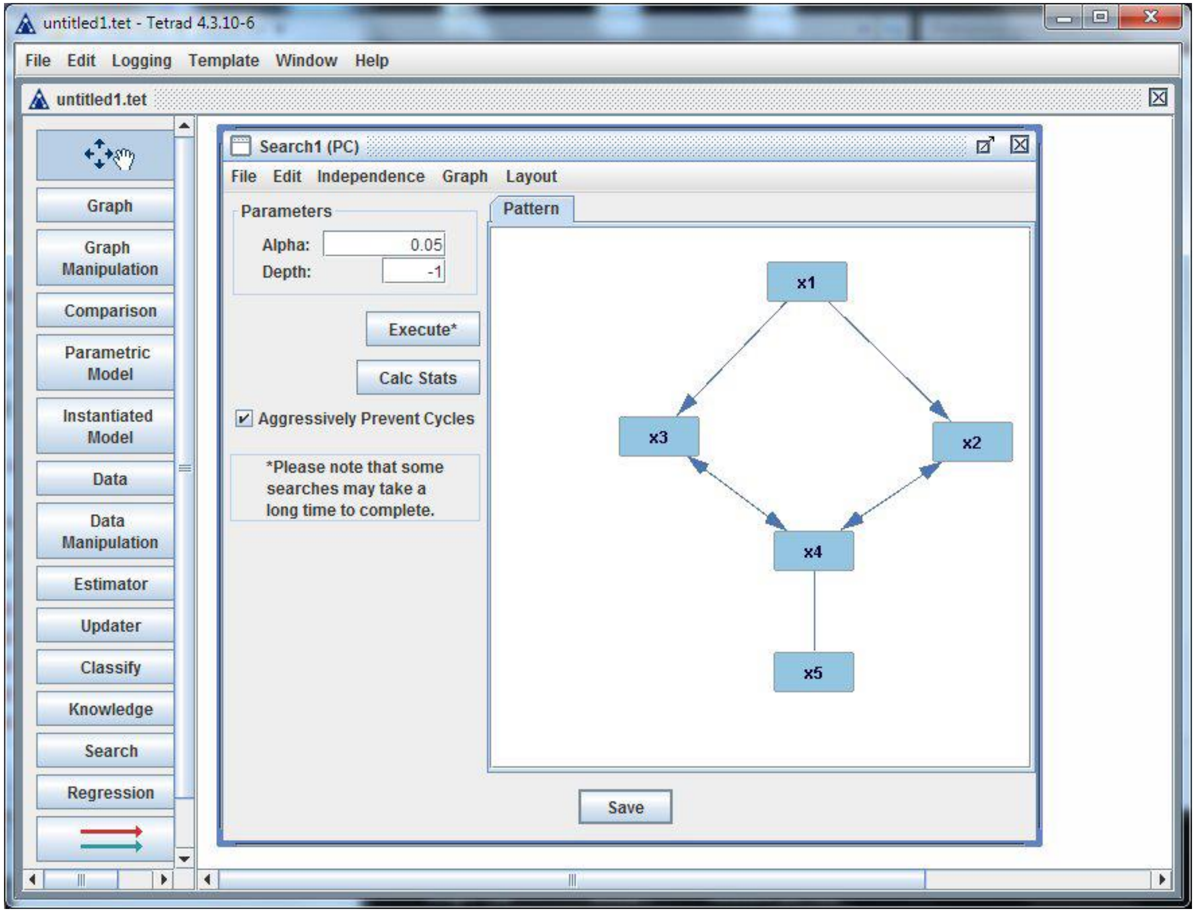
- We seek for particular representation of a true network – latent variable is direct, common cause of two observed variables (always exists)
- More complex algorithm, estimating MAG (Maximal Ancestral Graph)
- We may obtain more types of connections:
  - A *→ B
  - A → B that may mean A *→ B or A ← latent → B
  - A ←→ B that may mean A ← latent → B (spurious correlation)
  - A – B that may mean A *→ B or A ←* B or A ← latent → B

Examples of implementation

- Eshghi, Haughton, Topi. "Determinants of customer loyalty in the wireless telecommunications industry", *Telecommunications Policy*, 2007
- Bessler, Loper. "Economic Development: Evidence from Directed Acyclic Graphs", *The Manchester School*, 2001
- Rettenmaier, Wang. „What determines health: a causal analysis using county level data”, *Eur J Health Econ*, 2012

Software – Tetrad

- Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA
- http://www.phil.cmu.edu/projects/tetrad/current.html



Software – R

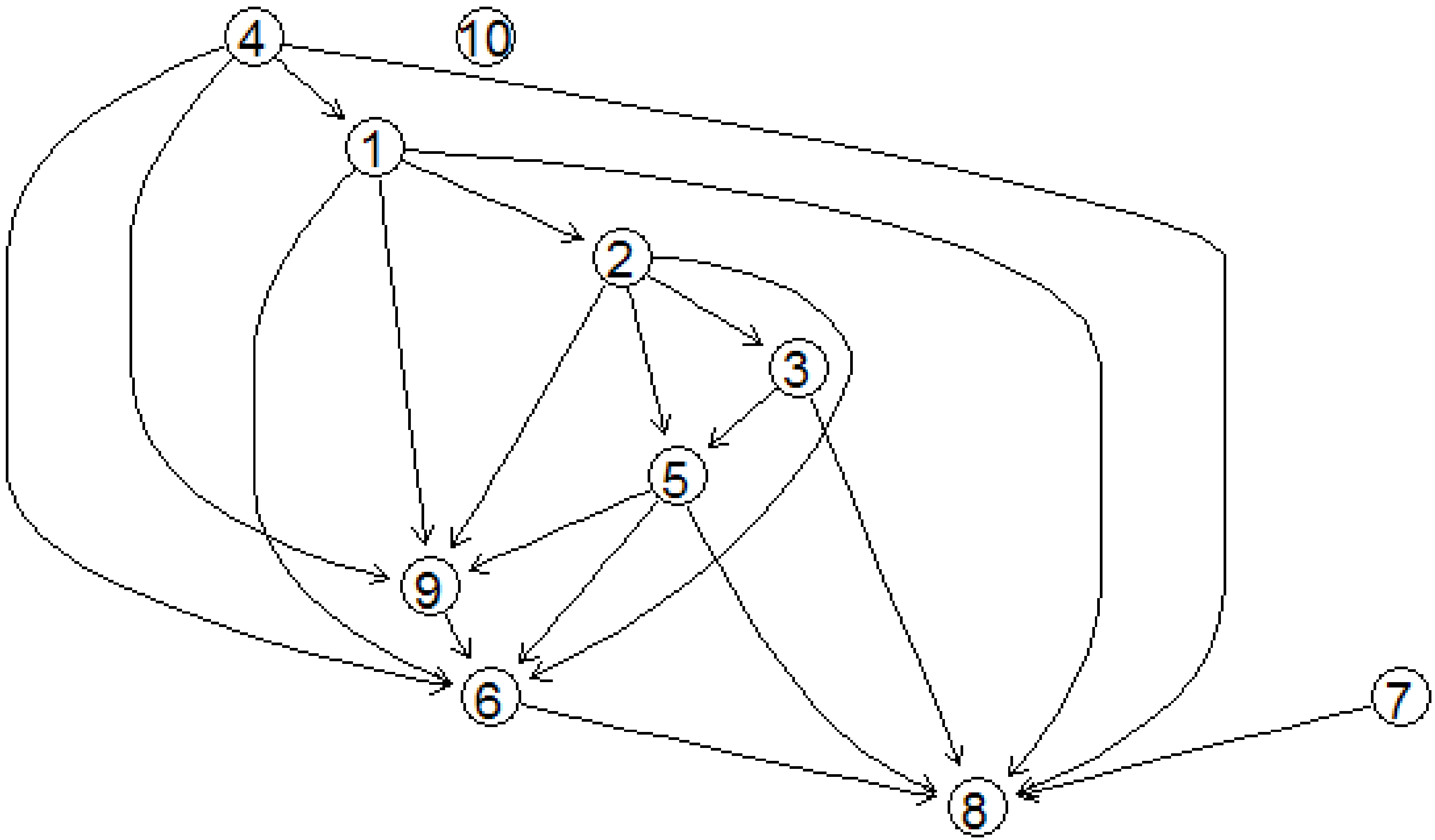
- R, library(pcalg)
- Data – Diagnoza Społeczna (binary, yet may be different)
- Exemplary code

```
library(pcalg)

dane <- read.csv(…)
nVar <- ncol(dane)

suffStat <- list(dm=dane,adaptDF=F)
indepTest <- binCItest

pc.fit <- pc(suffStat, indepTest, p = nVar, alpha = 0.01)
plot(pc.fit, main = "")
```

Data

- Financial problems (1)
- Health problems (2)
- Hierarchy of values – pleasure vs. helping others (3)
- Number of close friends (4)
- Goods are important for me (5)
- Can you trust other people (6)
- Can you trust your family (7)
- Can you trust your neighbours (8)
- Doing sports (9)
- Reading books (10)

Exemplary results



Exemplary insights

- (algorithm oriented all edges)
- Reading books do not cause, or are result of other variables
- Number of friends (social capital) impacts
  - Financial problems
  - Trust towards people/neighbours
  - Doing sports
- Financial problems → health problems
- Health problems → Hierarchy of values and importance of possession of goods
- Financial problems → Doing sports
- …

Impact strength analysis

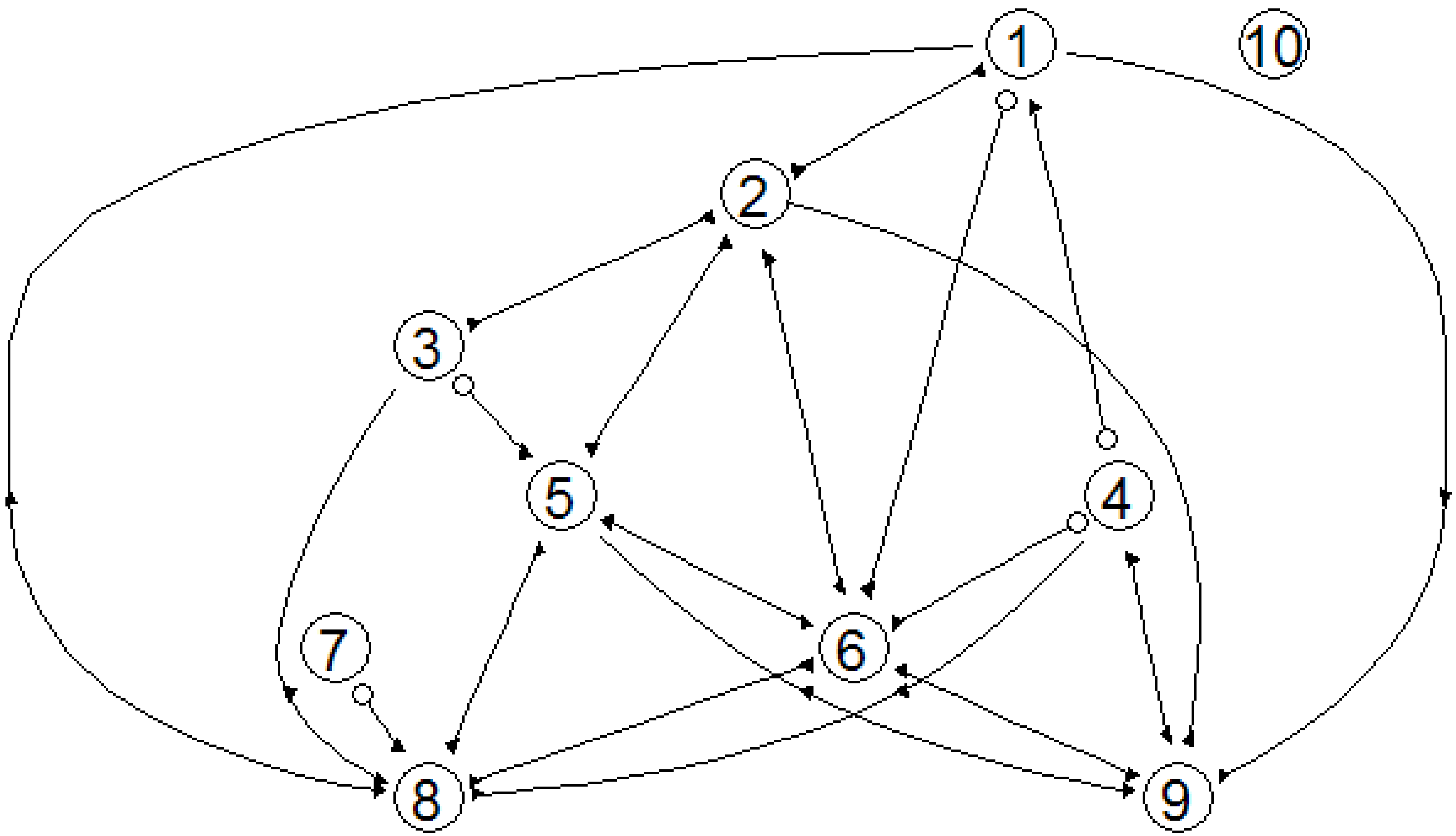
- How do financial problems impact health problems?

```
ida(x.pos=1, y.pos=2, cov(daneOryg),
pc.fit@graph, ...)
```

- Possible results as a set of values (due to equivalency class)
- We have single value: 0,15696
  - E(Y | do(X=1)) – E(Y | do(X=0))
- Therefore financial problems increase risk of health problems by 15,7 pp.

Result with latent variables

- Identical structure, many common causes, less informative directions



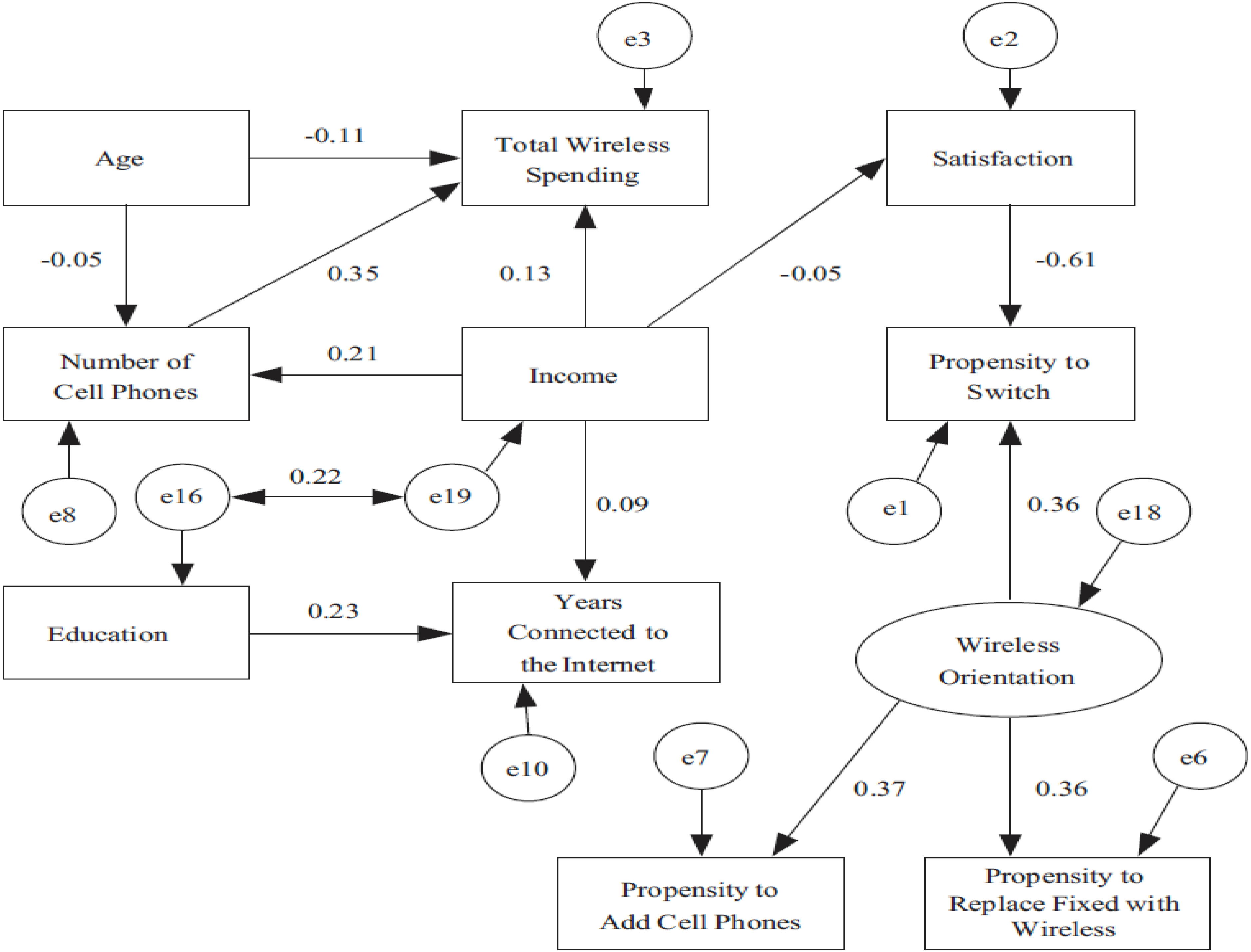
Examples

- Eshghi, Haughton, Topi „Determinants of customer loyalty in the wireless telecommunications industry”, Telecommunications Policy, 2007

- Data:
  - 3205 clients, telephone survey

Variable	Measurement/Description
Propensity to switch*	Probability (0–100%) to switch wireless providers within the next 12 months
Demographics	
Age*	Originally categorical 1–11 (under 18, 18–24, 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59, 60–64, 65 and older); categories replaced with the midpoints for the SEM
Education*	Originally categorical 1–5 (some high school, HS graduate, some college, college graduate, post graduate); categories replaced with 10, 12, 14, 16, and 18 for the SEM
Income*	Categorical 1–7, 88 (less than \$25,000; \$25,000-\$49,000; \$50,000-\$74,999; \$75,000-\$99,000; \$100,000-\$149,000; \$150,000—\$500,000; over \$500,000); categories replaced with 12.5, 37.5, 62.5, 87.5, 125, 325, and 750) for the SEM
Gender	1 = male, 0 = female
Technology intensity	
Owns a home computer?	Yes/no
Time on the Internet per week	Originally categorical 1–5 (0–5, 6–10, 11–15, 16–20, 21–); categories replaced with category midpoints (3, 8, 13, 18, 25) for the SEM
Years connected to the Internet?*	Originally categorical 1–4 (less than a year, 1–2 years, 3–5 years, 6– years); categories replaced with midpoints (1, 2, 4, 8) for the SEM
Level of higher technology involvement	
Number of cell phones*	Numeric
Propensity to add cell phones*	0–10
Total wireless spending*	In dollars
Propensity to replace fixed with wireless*	0–10
Percentage of long distance on wireless	0–100
Percent previously on fixed now on wireless	0–100
Ideal communications service provider characteristics	Respondents were asked to allocate a total of 100 points between the following 15 criteria of communications service providers to indicate how important each of these criteria are
Access	0–100
Clarity	0–100
Community-minded	0–100
Customization	0–100
Employee empowerment	0–100
Expertise	0–100
Freedom	0–100
Innovation	0–100
Integration	0–100
Market leader	0–100
Nimble	0–100
Reliable	0–100
Timeliness	0–100
Simplicity	0–100
Trustworthy	0–100
Satisfaction*	0–10

Note: The variables marked with an asterisk were included in the final structural equation model.


e3
e2
Age
-0.11
Total Wireless Spending
Satisfaction
-0.05
0.35
0.13
-0.05
-0.61
Number of Cell Phones
0.21
Income
Propensity to Switch
e8
e16
0.22
e19
0.09
e1
0.36
e18
Education
0.23
Years Connected to the Internet
Wireless Orientation
e10
e7
0.37
0.36
e6
Propensity to Add Cell Phones
Propensity to Replace Fixed with Wireless

D-separation

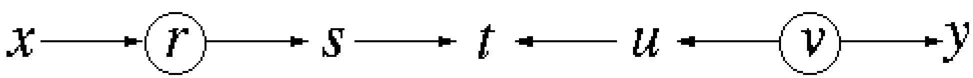
- Proper calculation of impact strength requires blocking non-causal Paths
- Path is any sequence of vertices connected by edges, regardless of directions
- Causal path is a directed Path from A to B
- Collider is a vertex with in-edges and without out-edges
- Path with collider is blocked
- In order to block Path without colliders one must use conditioning
- In order to open Path with collider one must condition on collider

D-separation

Rule 1: *x* and *y* are *d*-connected if there is an unblocked path between them.

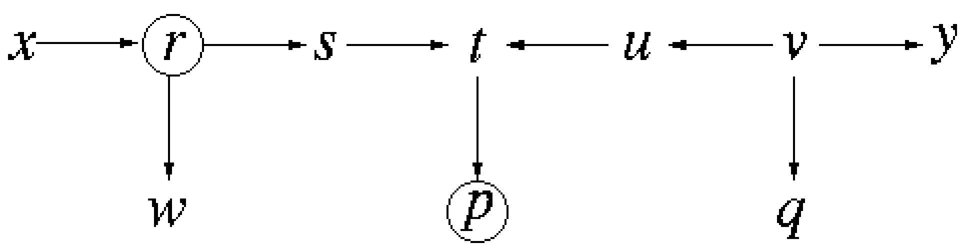
R and V are d-separated
X and T are d-connected

Rule 2: *x* and *y* are *d*-connected, conditioned on a set *Z* of nodes, if there is a collider-tree path between *x* and *y* that traverses no member of *Z*. If no such path exists, we say that *x* and *y* are *d*-separated by *Z*, We also say then that every path between *x* and *y* is "blocked" by *Z*.



X and S are d-separated given {R,V}
S and T are d-connected

Rule 3: If a collider is a member of the conditioning set *Z*, or has a descendant in *Z*, then it no longer blocks any path that traces this collider.



X and U are d-separated given {R,P}
S and Y are d-connected given {R,P}

Summary

- It is possible to detect causal relationships in observational data
- It requires assumptions on graph structure
- Mechanisms of causal inference are intuitive
- Introduction of latent variables results in more difficult inference

References

- D. Colombo et al. „Learning High-Dimensional Directed Acyclic Graphs with Latent and Selection Variables", *The Annals of Statistics*, 2012
- J. Pearl, *Causality. Models, Reasoning, and Inference*, 2009
- J. Pearl, Causal inference in statistics: An overview. *Statistics Surveys*, 2009
- Woodward J. *Making Things Happen. A Theory of Causal Explanation*. 2003