CWTS Meaningful metrics

Causal diagrams

Vincent Traag

WTMC method track – Computational methods & modelling

11 February 2021

Universiteit Leiden

1. Observing $\neq$ doing

2. *d*-separation simplified

Observing $\neq$ doing

Predicting

- Predict future citations

CWTS Meaningful metrics

Predicting

- Predict future citations
  - Tweets predict future citations

Predicting

- Predict future citations
  - Tweets predict future citations
- Predict accept/reject decisions

CWTS Meaningful metrics

Predicting

- Predict future citations
  - Tweets predict future citations
- Predict accept/reject decisions
  - Early publication record predicts acceptance

CWTS Meaningful metrics

Predicting

- Predict future citations
  - Tweets predict future citations
- Predict accept/reject decisions
  - Early publication record predicts acceptance
- Predict future career paths

Predicting

- Predict future citations
  - Tweets predict future citations
- Predict accept/reject decisions
  - Early publication record predicts acceptance
- Predict future career paths
  - Status of PhD supervisor predicts future career

CWTS Meaningful metrics

Predicting

- Predict future citations
  - Tweets predict future citations
- Predict accept/reject decisions
  - Early publication record predicts acceptance
- Predict future career paths
  - Status of PhD supervisor predicts future career
- Predict breakthrough papers

CWTS Meaningful metrics

Predicting

- Predict future citations
  - Tweets predict future citations
- Predict accept/reject decisions
  - Early publication record predicts acceptance
- Predict future career paths
  - Status of PhD supervisor predicts future career
- Predict breakthrough papers
  - Multidisciplinary citations predict breakthrough papers

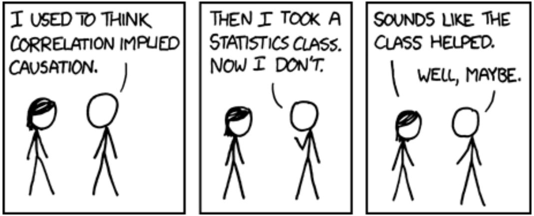
Predicting

- Predict future citations
  - Tweets predict future citations
- Predict accept/reject decisions
  - Early publication record predicts acceptance
- Predict future career paths
  - Status of PhD supervisor predicts future career
- Predict breakthrough papers
  - Multidisciplinary citations predict breakthrough papers

Observations $\rightarrow$ prediction

CWTS Meaningful metrics

Observation



CWTS
Meaningful metrics

Observing vs doing

Prediction

Observing x may make y more likely to be observed.

CWTS
Meaningful metrics

Observing vs doing

Prediction

Observing x may make y more likely to be observed.

Intervention

Doing x may make y more likely to be observed.

Observation vs doing

- Predict future citations
  - Tweets predict future citations

CWTS Meaningful metrics

Observation vs doing

- Predict future citations
  - Tweets predict future citations
  - Does tweeting your paper increase citations?

Observation vs doing

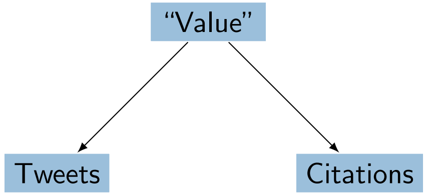
- Predict future citations
  - Tweets predict future citations
  - Does tweeting your paper increase citations?

$$\text{Tweets} \longrightarrow \text{Citations}$$

CWTS Meaningful metrics

Observation vs doing

- Predict future citations
  - Tweets predict future citations
  - Does tweeting your paper increase citations?



Observation vs doing

- Predict accept/reject decisions
  - Early publication record predicts acceptance

Observation vs doing

- Predict accept/reject decisions
  - Early publication record predicts acceptance
  - Does increasing the publication record make an “accept” more likely?

Observation vs doing

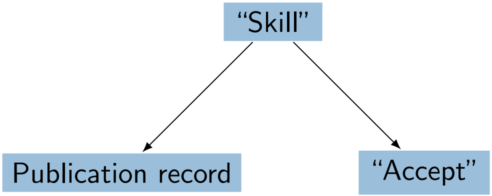
- Predict accept/reject decisions
  - Early publication record predicts acceptance
  - Does increasing the publication record make an “accept” more likely?

Publication record → “Accept”

CWTS
Meaningful metrics

Observation vs doing

- Predict accept/reject decisions
  - Early publication record predicts acceptance
  - Does increasing the publication record make an "accept" more likely?



CWTS Meaningful metrics

Causal inference

Causality

- Causality is about *doing* something, about (hypothetical) interventions.
- Important for policy implications:
  - If you observe x and y are associated, but they are *not* causally related, then *doing* x does not change y, and x should not be advocated as a policy to achieve a change in y.

CWTS Meaningful metrics

Causal inference

Causality

- Causality is about *doing* something, about (hypothetical) interventions.
- Important for policy implications:
  - If you observe x and y are associated, but they are *not* causally related, then *doing* x does not change y, and x should not be advocated as a policy to achieve a change in y.

How to discern association from causation?

CWTS
Meaningful metrics

Causal inference

Causality

- Causality is about *doing* something, about (hypothetical) interventions.
- Important for policy implications:
  - If you observe x and y are associated, but they are *not* causally related, then *doing* x does not change y, and x should not be advocated as a policy to achieve a change in y.

How to discern association from causation?

Causal inference

1. Assume a causal model (background knowledge, earlier research, ...).

CWTS Meaningful metrics

Causal inference

Causality

- Causality is about *doing* something, about (hypothetical) interventions.
- Important for policy implications:
  - If you observe x and y are associated, but they are *not* causally related, then *doing* x does not change y, and x should not be advocated as a policy to achieve a change in y.

How to discern association from causation?

Causal inference

1. Assume a causal model (background knowledge, earlier research, ...).
2. Given causal model, under some conditions associations may reveal *causal* effects.
   - Key to this is understanding d-separation.

Causal inference

Causality

- Causality is about *doing* something, about (hypothetical) interventions.
- Important for policy implications:
  - If you observe x and y are associated, but they are *not* causally related, then *doing* x does not change y, and x should not be advocated as a policy to achieve a change in y.

How to discern association from causation?

Causal inference

1. Assume a causal model (background knowledge, earlier research, ...).
2. Given causal model, under some conditions associations may reveal *causal* effects.
   - Key to this is understanding d-separation.
3. Test whether implications of causal model hold in observations.
   - Different causal models may have same observational implications.

CWTS
Meaningful metrics

d-separation simplified

Universiteit Leiden

Two variables x and y are associated if between them there is an open undirected path

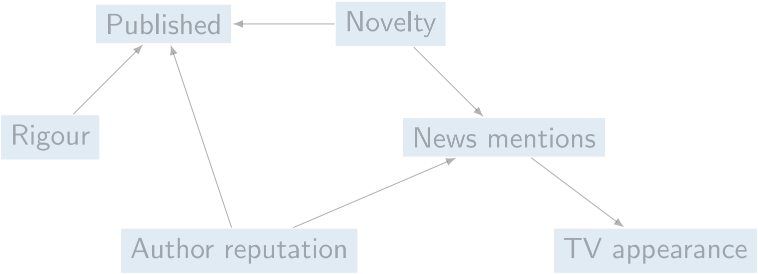
Two variables x and y are associated if between them there is an open undirected path

Two variables x and y are associated if between them there is an open undirected path

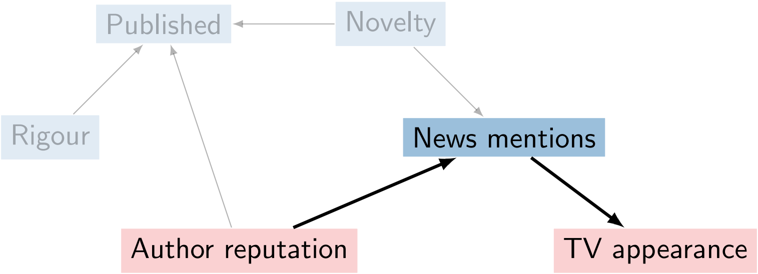
Two variables x and y are associated if between them there is an open undirected path

Technical term	Meaning	Implication
d-connected	there is an open undirected path	x and y are associated
d-separated	there is no open undirected path	x and y are not associated

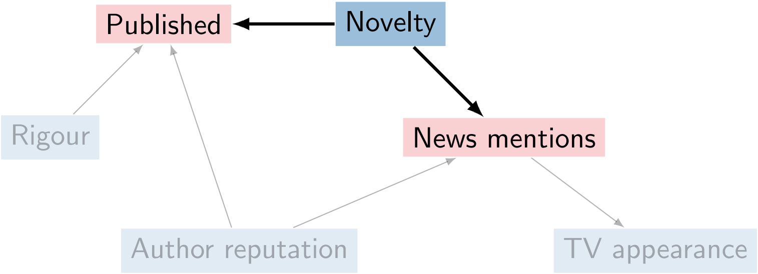
d-separation – Undirected paths



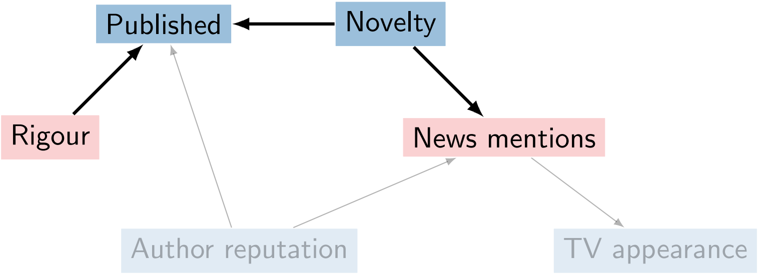
d-separation – Undirected paths



d-separation – Undirected paths



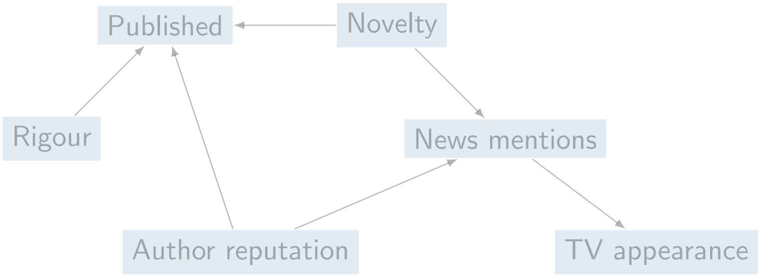
d-separation – Undirected paths



An undirected path is open if *all* nodes on the path are open .

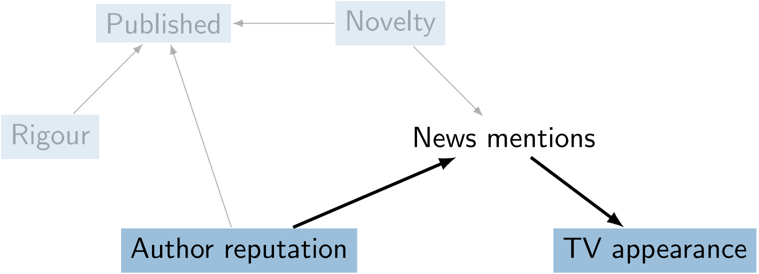
An undirected path is open if *all* nodes on the path are open.

d-separation – Mediators, confounders & colliders



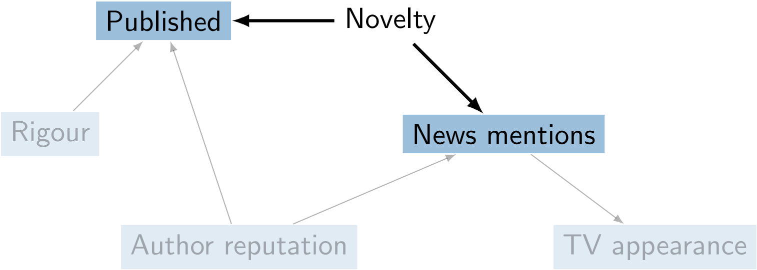
Mediator (causation) – Open

$$\cdots \longrightarrow z \longrightarrow \cdots$$



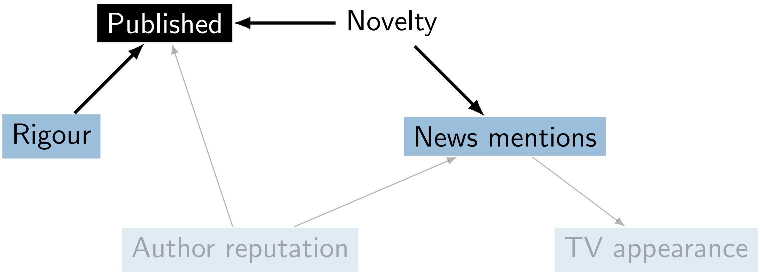
Confounder – Open

$$\cdots \longleftarrow z \longrightarrow \cdots$$



CWTS
Meaningful metrics

Collider – Closed



Conditioning on variables

Many possible ways to “condition” on a variable:

- Include only subset of cases in analysis.
- Perform analyses separately on different subsets of data.
- “Normalise” indicators (e.g field-normalise).
- Include variable in regression.

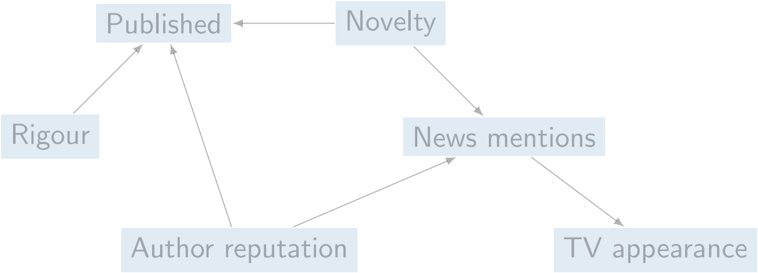
CWTS
Meaningful metrics

d-separation – Open nodes & conditioning

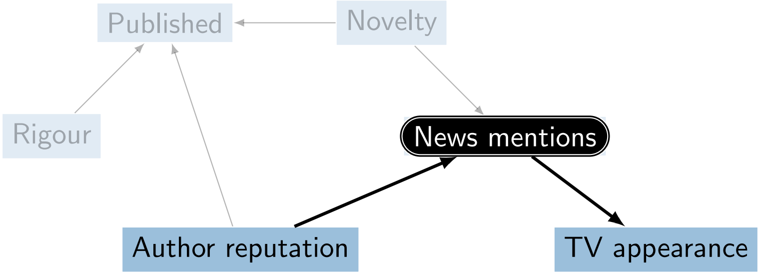
Conditioning on a variable “flips” the node on a path
from open □ to closed ■ and vice-versa.

	Unconditioned	Conditioned
Mediator	$\cdots \rightarrow z \rightarrow \cdots$	$\cdots \rightarrow z \rightarrow \cdots$
Confounder	$\cdots \leftarrow z \rightarrow \cdots$	$\cdots \leftarrow z \rightarrow \cdots$
Collider	$\cdots \rightarrow \boxed{z} \leftarrow \cdots$	$\cdots \rightarrow z \leftarrow \cdots$

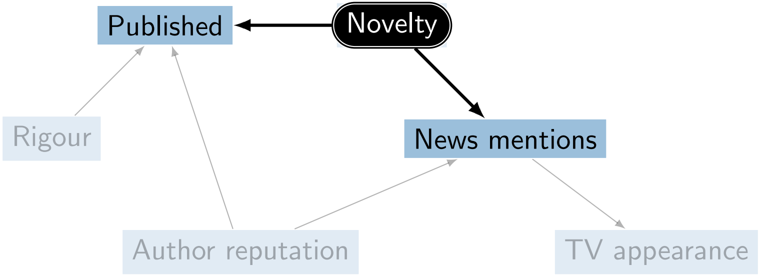
d-separation – Conditioning



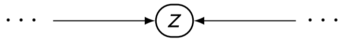
Condition on mediator – Closed

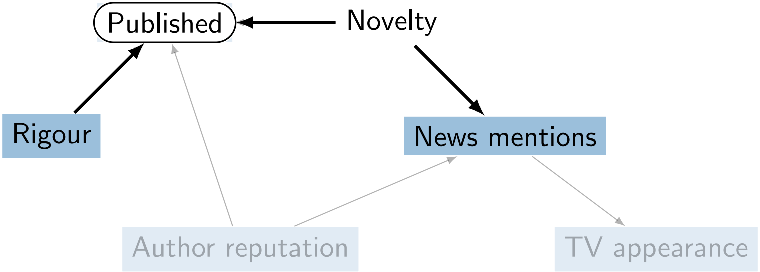


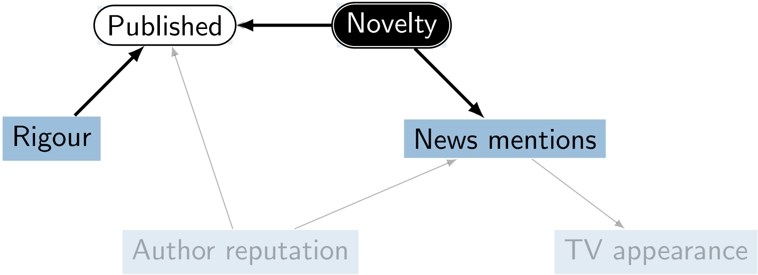
Condition on confounder – Closed



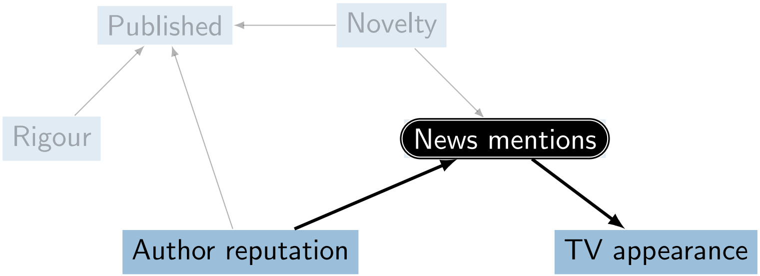
Condition on collider – Open





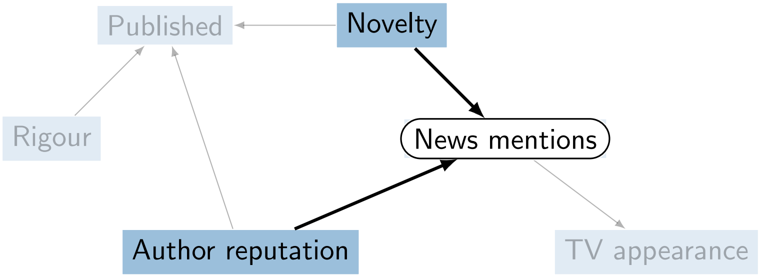

Published
Novelty
Rigour
News mentions
Author reputation
TV appearance

Conditioning may close one path but open another.



d-separation – Conditioning

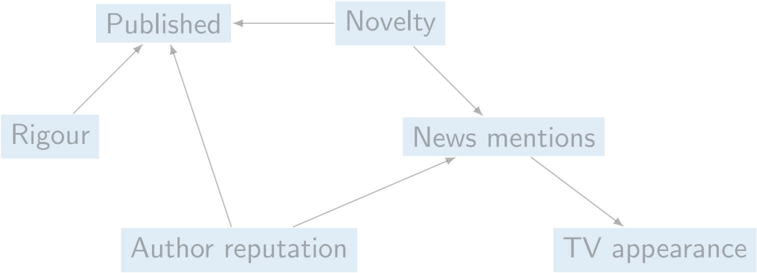
Conditioning may close one path but open another.



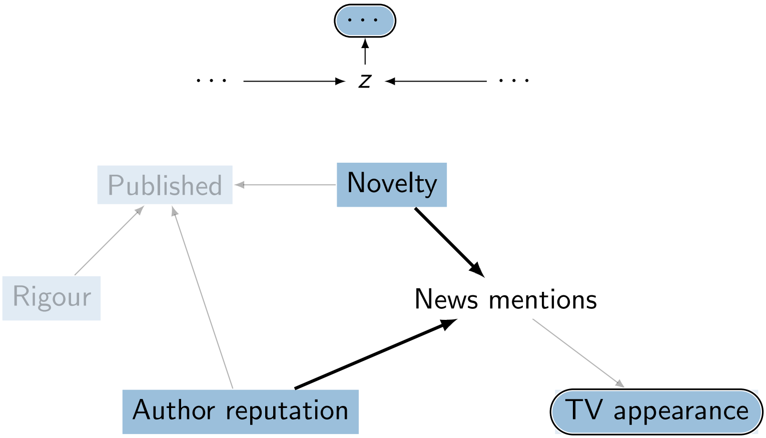
Conditioning on descendant of collider opens it.

	Unconditioned	Conditioned	Descendant conditioning
Mediator	$\cdots \rightarrow Z \rightarrow \cdots$	$\cdots \rightarrow Z \rightarrow \cdots$	
Confounder	$\cdots \leftarrow Z \rightarrow \cdots$	$\cdots \leftarrow Z \rightarrow \cdots$	
Collider	$\cdots \rightarrow \boxed{Z} \leftarrow \cdots$	$\cdots \rightarrow Z \leftarrow \cdots$	$\cdots \rightarrow Z \leftarrow \cdots$, $Z \rightarrow \cdots$

d-separation – Conditioning on descendants



Condition on descendant of collider – Opens collider



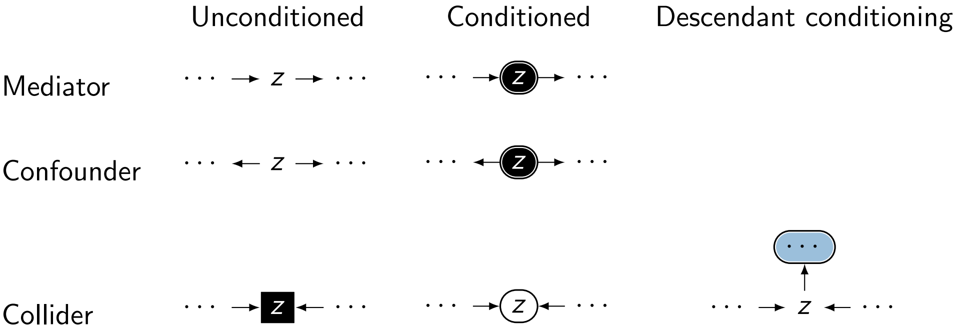
1. x and y are associated if there is an open undirected path between them.

d-separation summary

1. x and y are associated if there is an open undirected path between them.
2. An undirected path is open if all nodes are open.

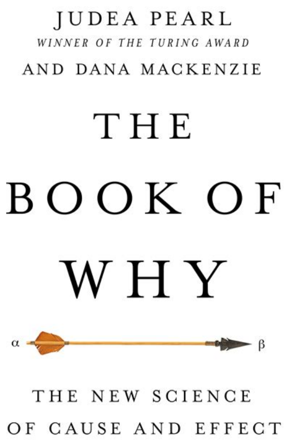
CWTS Meaningful metrics

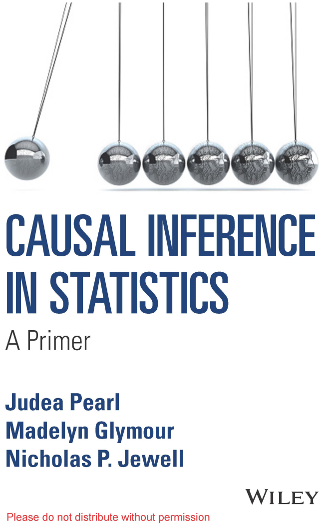
1. x and y are associated if there is an open undirected path between them.
2. An undirected path is open if all nodes are open.
3. A node on a path is open □ or closed ■ in the following situations
   - Conditioning on a variable "flips" nodes from open to closed and vice-versa.



CWTS Meaningful metrics

Further reading





CWTS
Meaningful metrics