Graphical Models

Causality

Marc Maier
April 28, 2011

k n o w l e d g e
d i s c o v e r y
l a b o r a t o r y

Topics

Causality described

Causal assumptions

Causal discovery

Causation

*"The paradigmatic assertion in causal relationships
is that manipulation of a cause
will result in the manipulation of an effect...*

Causation implies that
by varying one factor,
I can make another vary."

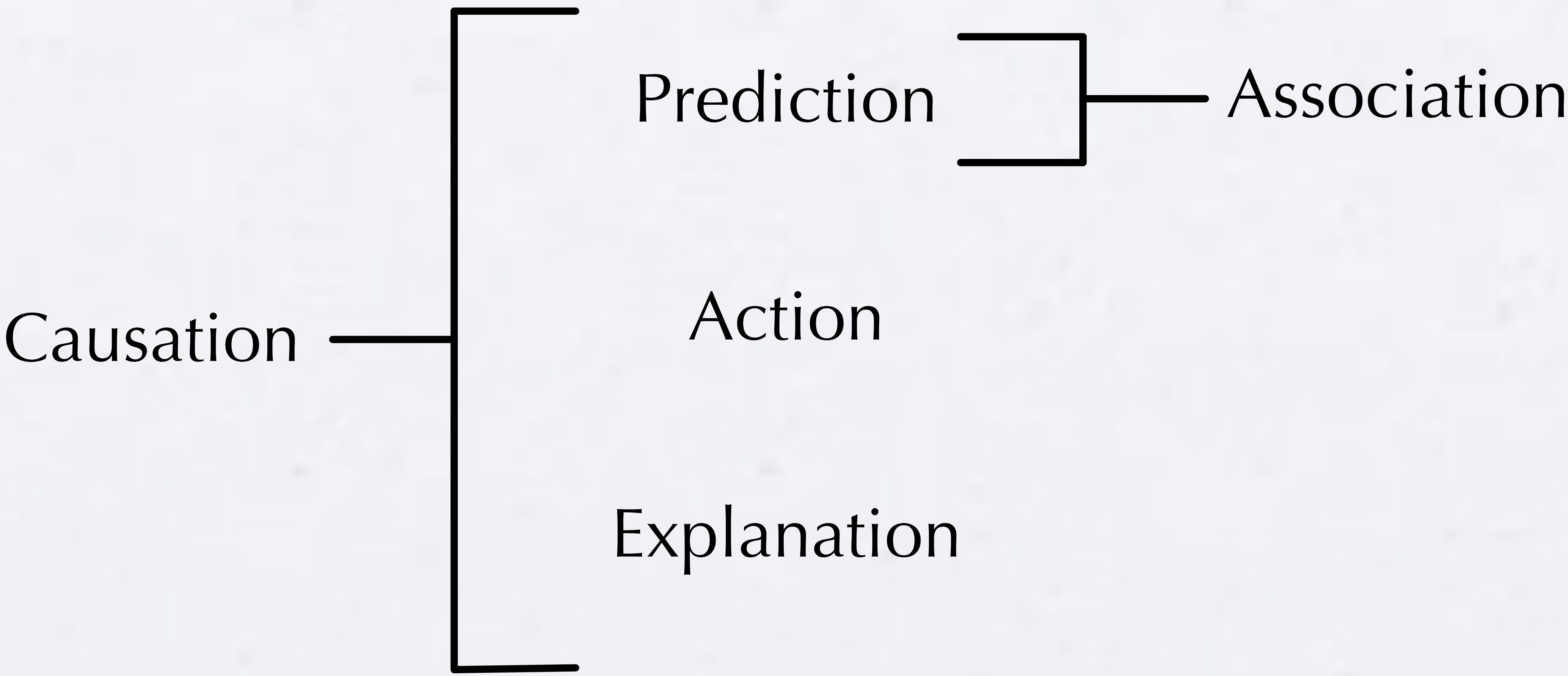
(Cook & Campbell 1979)

Probabilistic causation:

$$\exists x, x' \; P(Y = y | do(X = x)) > P(Y = y | do(X = x'))$$

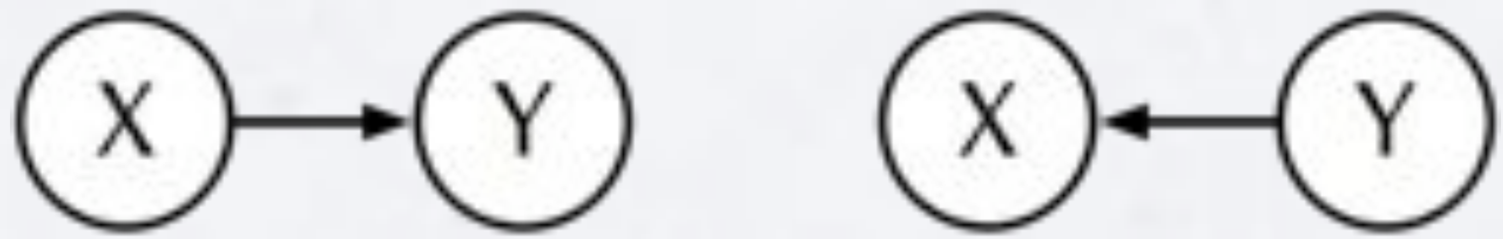
(Pearl 2000)

Associational vs. Causal Models

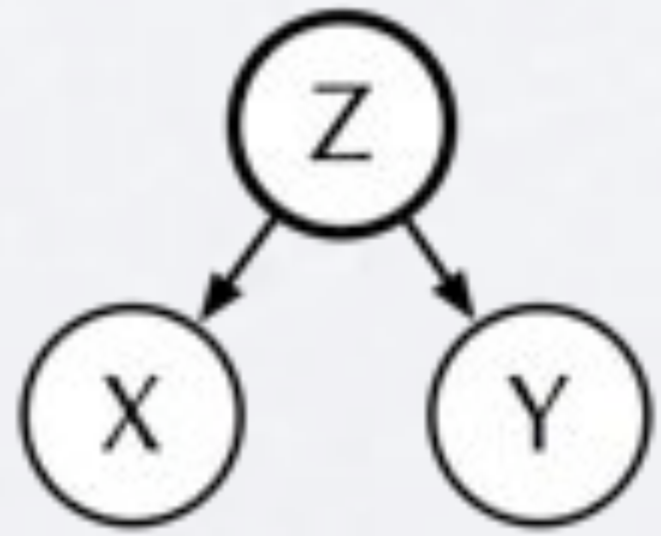


Association *underdetermines* Causation

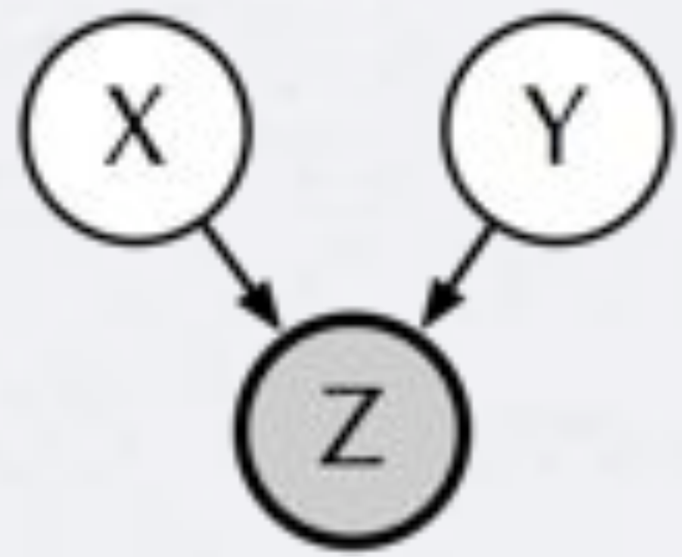
Direction



Common Causes



Common Effects



Propositional data representation

- Independent
- Identically distributed

movie	budget	gross	genre	year
Goodfellas	25M	47M	crime	1990
My Cousin Vinny	11M	64M	comedy	1992
...	...	...	...	...
Clue	15M	15M	comedy	1985

Directed Acyclic Graph

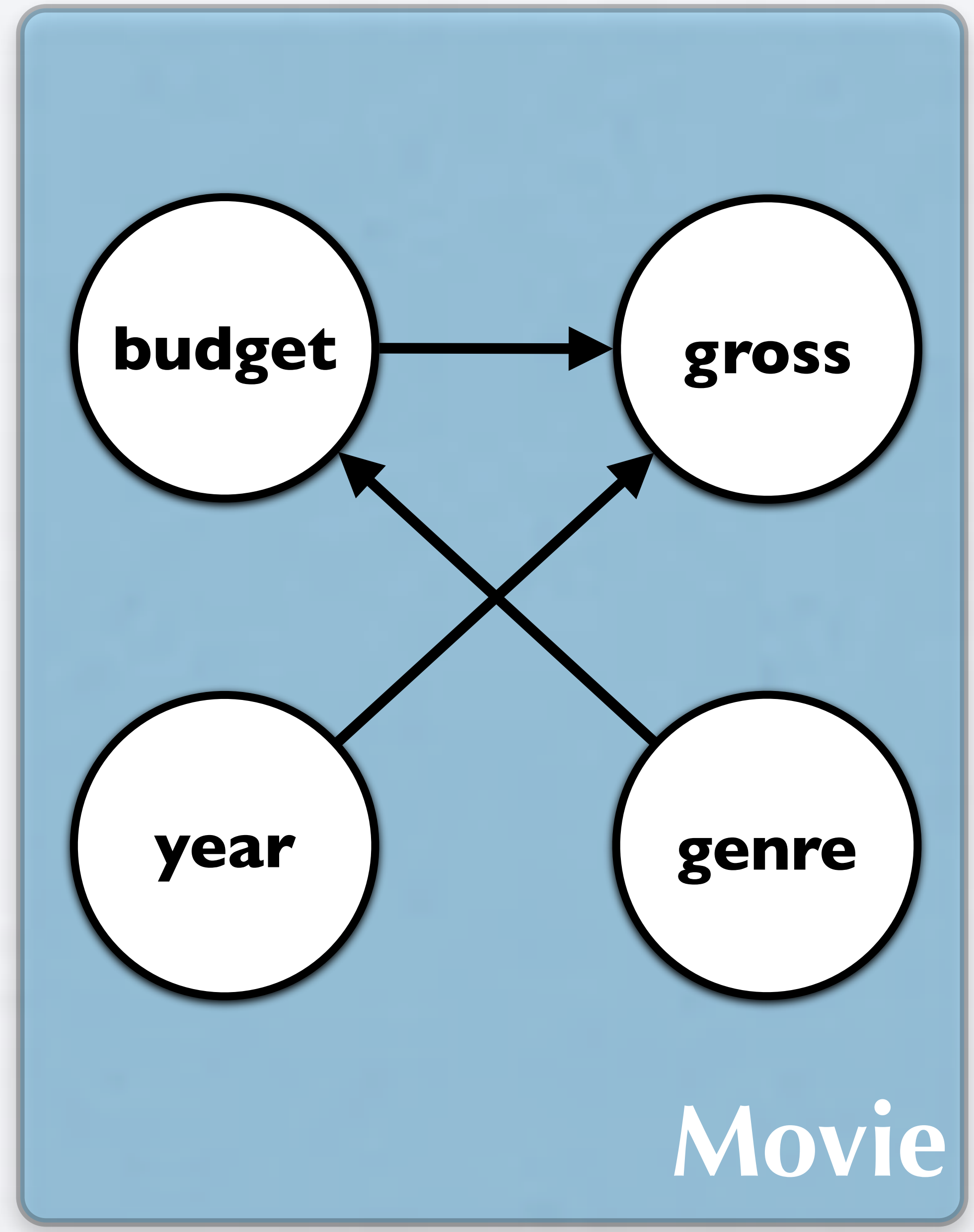
- Random variables

$$V = \{\text{year}, \text{genre}, \text{budget}, \text{gross}\}$$

- Conditional independencies

e.g., $\text{genre} \perp\!\!\!\perp \text{gross} \mid \text{budget}$

- Joint probability distribution



$$p(V) = p(\text{year})p(\text{genre})p(\text{budget}|\text{genre})p(\text{gross}|\text{budget}, \text{year})$$

(Pearl 1988; 2000)

Structure learning paradigms

Search-and-score: Perform *global* search across model space, select one with highest likelihood

Constraint-based: Run *local* tests of independence to create constraints on space of possible models

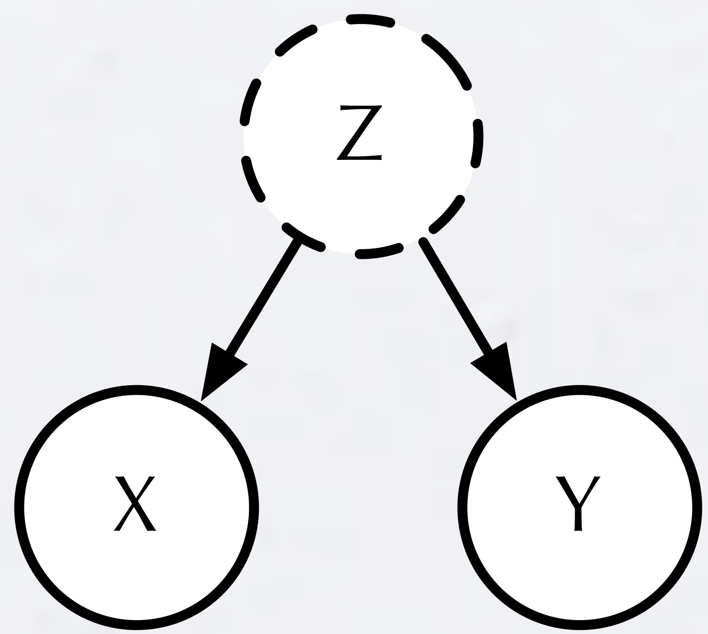
Structure learning paradigms

	Pros	Cons
S&S	• Approximates joint distribution • Smooth/Bayesian (not prone to unstable errors)	• Computationally intensive (NP-hard) • No theoretical guarantees • May choose single model from equivalence class
CB	• Separates structure learning from parameter estimation • Directly learns conditional independence relations • Provably correct • Can be efficient • Extensible to other new operations	• Individual errors may propagate

Causal Assumptions

Causal sufficiency

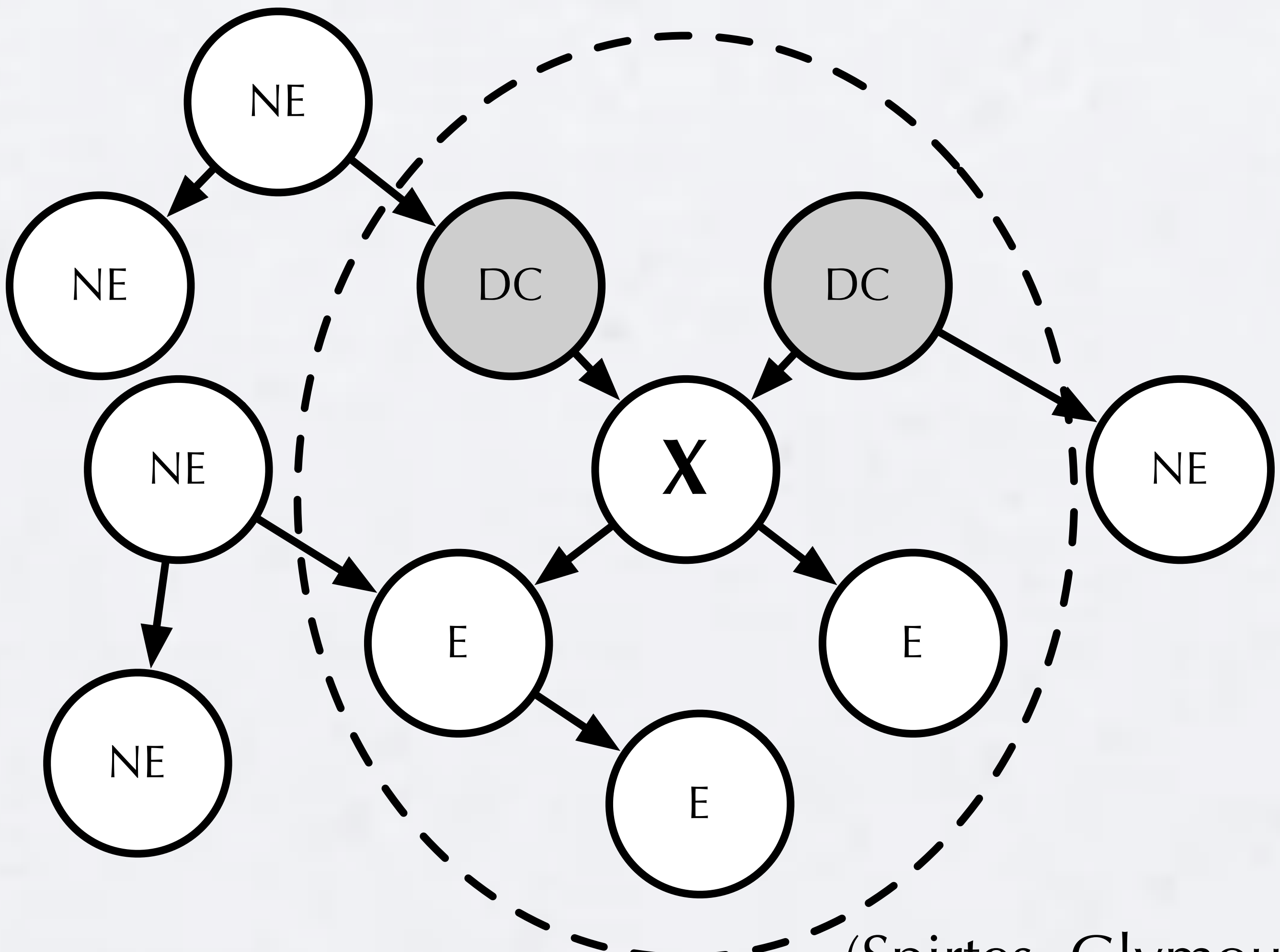
$\mathcal{V}$ is *causally sufficient* if and only if
for all potential causal dependencies $\langle X, Y \rangle \in \mathcal{V} \times \mathcal{V}$,
all common causes are measured and included in $\mathcal{V}$



(Spirtes, Glymour, Scheines 1993; 2001)

Causal Markov condition

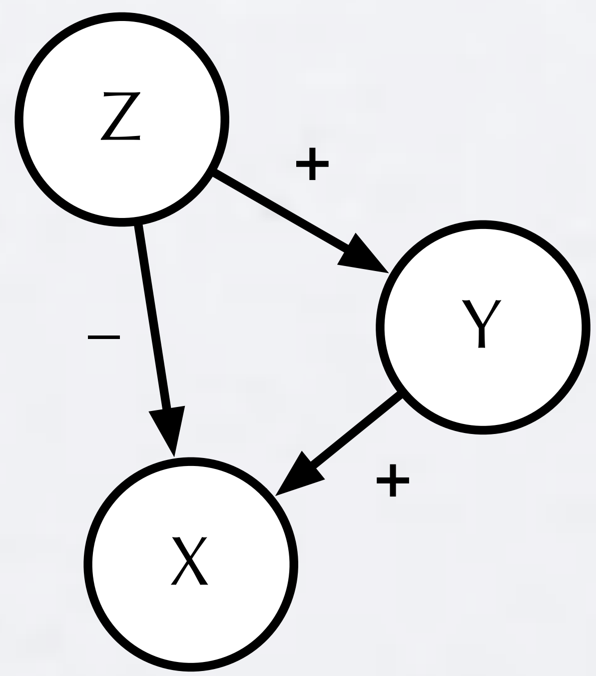
Given that $\mathcal{V}$ is causally sufficient,
$\mathcal{P}$ is *Markov* to $\mathcal{G}$ if and only if
each variable $X \in \mathcal{V}$ is conditionally independent
of its non-effects given its direct causes



(Spirtes, Glymour, Scheines 1993; 2001)

Faithfulness

$\mathcal{P}$ is *faithful* to $\mathcal{G}$ if and only if there exist no conditional independencies in $\mathcal{P}$ not entailed by the causal Markov condition on $\mathcal{G}$



(Spirtes, Glymour, Scheines 1993; 2001)

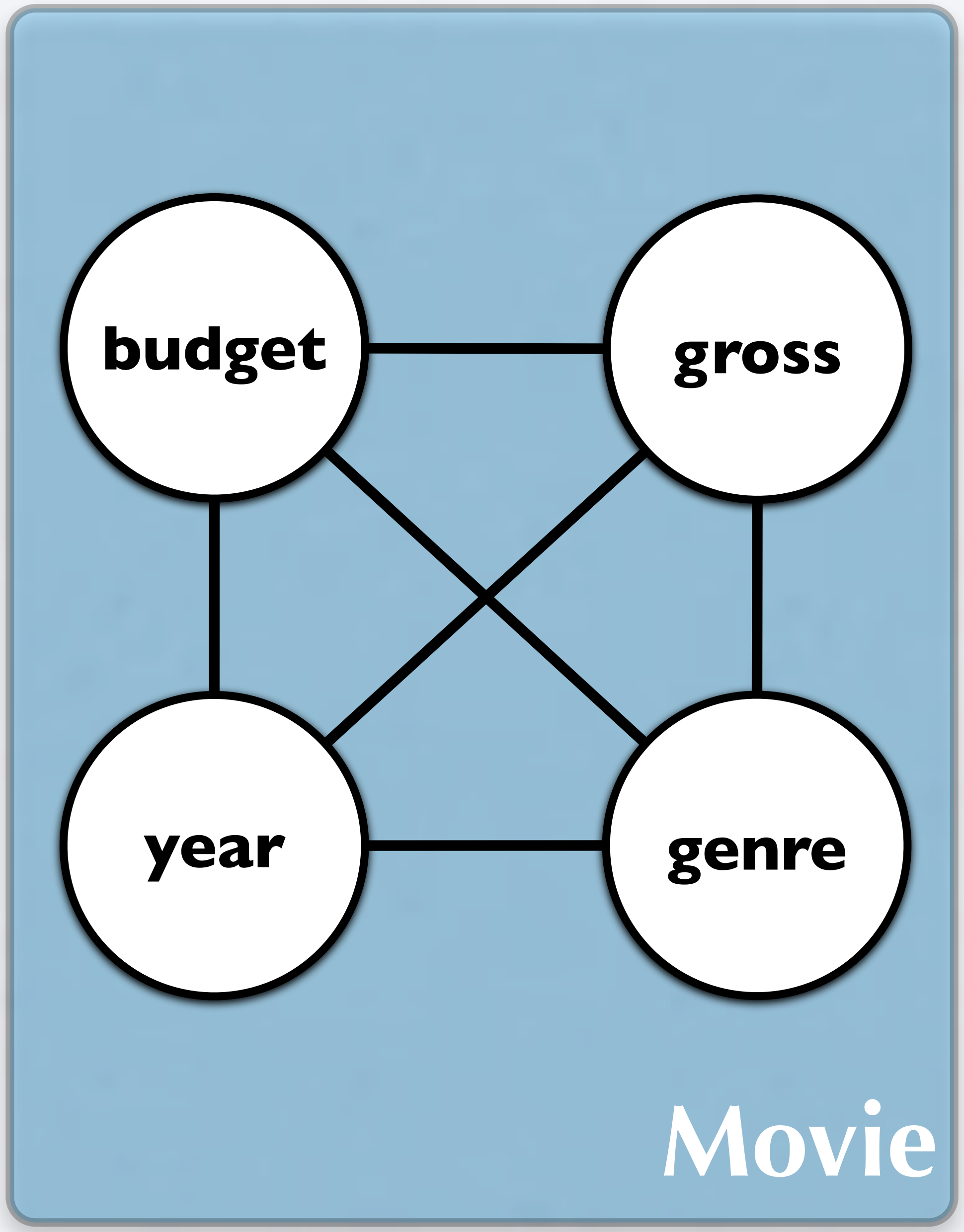
SGS

The SGS algorithm

Phase I

Skeleton identification

Determine set of conditional independencies among all variables



(Spirtes, Glymour, Scheines 1993; 2001)

The SGS algorithm

budget $\perp\!\!\!\perp$ genre | {}
budget $\perp\!\!\!\perp$ gross | {}
budget $\perp\!\!\!\perp$ year | {}
genre $\perp\!\!\!\perp$ budget | {}
genre $\perp\!\!\!\perp$ gross | {}
genre $\perp\!\!\!\perp$ year | {}
gross $\perp\!\!\!\perp$ budget | {}
gross $\perp\!\!\!\perp$ genre | {}
gross $\perp\!\!\!\perp$ year | {}
year $\perp\!\!\!\perp$ budget | {}
year $\perp\!\!\!\perp$ genre | {}
year $\perp\!\!\!\perp$ gross | {}
budget $\perp\!\!\!\perp$ genre | {gross}
budget $\perp\!\!\!\perp$ genre | {year}
budget $\perp\!\!\!\perp$ gross | {genre}
budget $\perp\!\!\!\perp$ gross | {year}
budget $\perp\!\!\!\perp$ year | {genre}
budget $\perp\!\!\!\perp$ year | {gross}
genre $\perp\!\!\!\perp$ budget | {gross}
genre $\perp\!\!\!\perp$ budget | {year}
genre $\perp\!\!\!\perp$ gross | {budget}
genre $\perp\!\!\!\perp$ gross | {year}
genre $\perp\!\!\!\perp$ year | {budget}
genre $\perp\!\!\!\perp$ year | {gross}
gross $\perp\!\!\!\perp$ budget | {genre}
gross $\perp\!\!\!\perp$ budget | {year}
gross $\perp\!\!\!\perp$ genre | {budget}
gross $\perp\!\!\!\perp$ genre | {year}
gross $\perp\!\!\!\perp$ year | {budget}
gross $\perp\!\!\!\perp$ year | {genre}
year $\perp\!\!\!\perp$ budget | {genre}
year $\perp\!\!\!\perp$ budget | {gross}
year $\perp\!\!\!\perp$ genre | {budget}
year $\perp\!\!\!\perp$ genre | {gross}
year $\perp\!\!\!\perp$ gross | {budget}
year $\perp\!\!\!\perp$ gross | {genre}
budget $\perp\!\!\!\perp$ genre | {gross, year}
budget $\perp\!\!\!\perp$ gross | {genre, year}
budget $\perp\!\!\!\perp$ year | {genre, gross}
genre $\perp\!\!\!\perp$ budget | {gross, year}
genre $\perp\!\!\!\perp$ gross | {budget, year}
genre $\perp\!\!\!\perp$ year | {budget, gross}
gross $\perp\!\!\!\perp$ budget | {genre, year}
gross $\perp\!\!\!\perp$ genre | {budget, year}
gross $\perp\!\!\!\perp$ year | {budget, genre}
year $\perp\!\!\!\perp$ budget | {genre, gross}
year $\perp\!\!\!\perp$ genre | {budget, gross}
year $\perp\!\!\!\perp$ gross | {budget, genre}

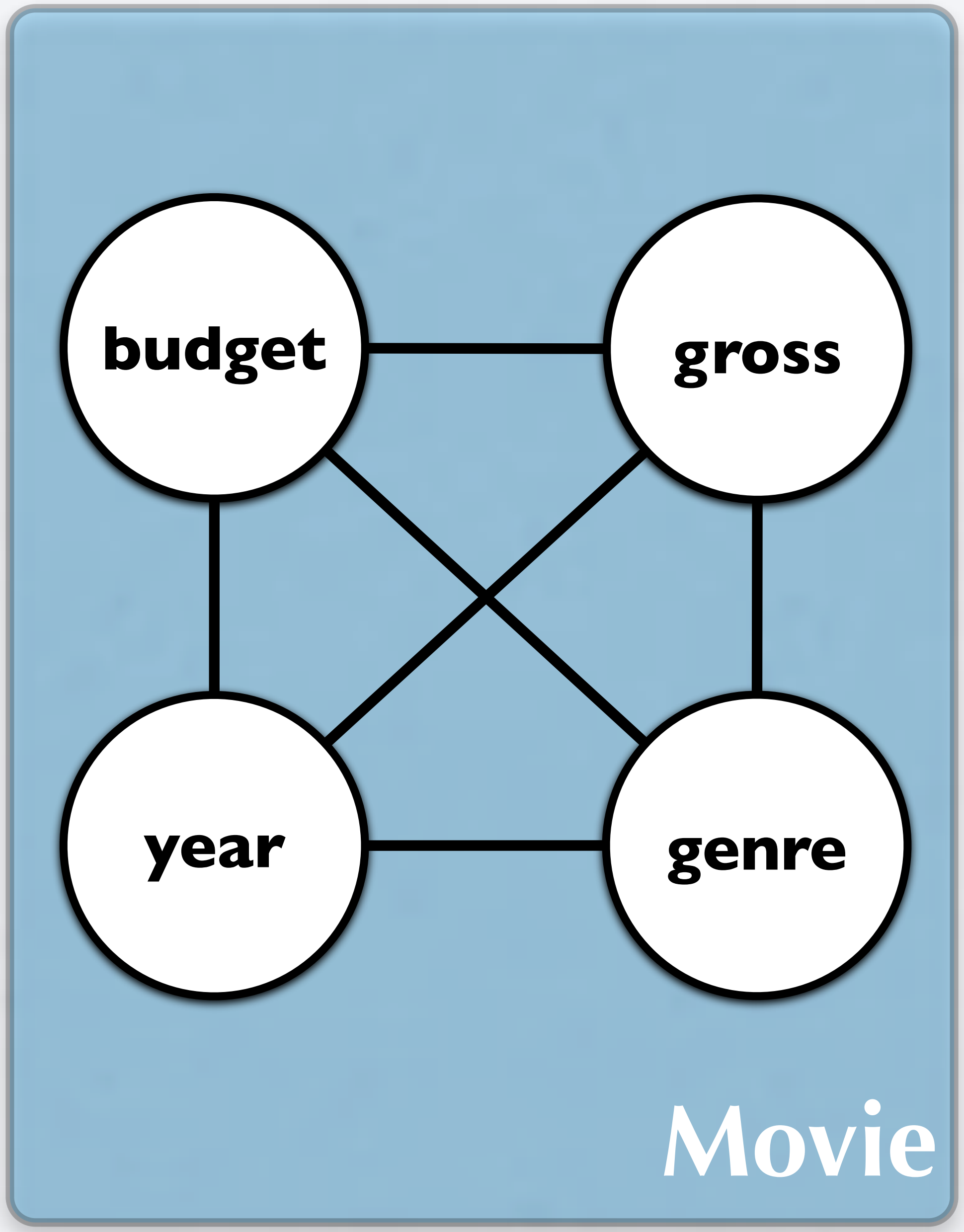
(Spirtes, Glymour, Scheines 1993; 2001)

The SGS algorithm

<u>Phase I</u>

Skeleton identification

Determine set of conditional independencies among all variables



$$\text{budget} \perp\!\!\!\perp \text{year} \ ?$$

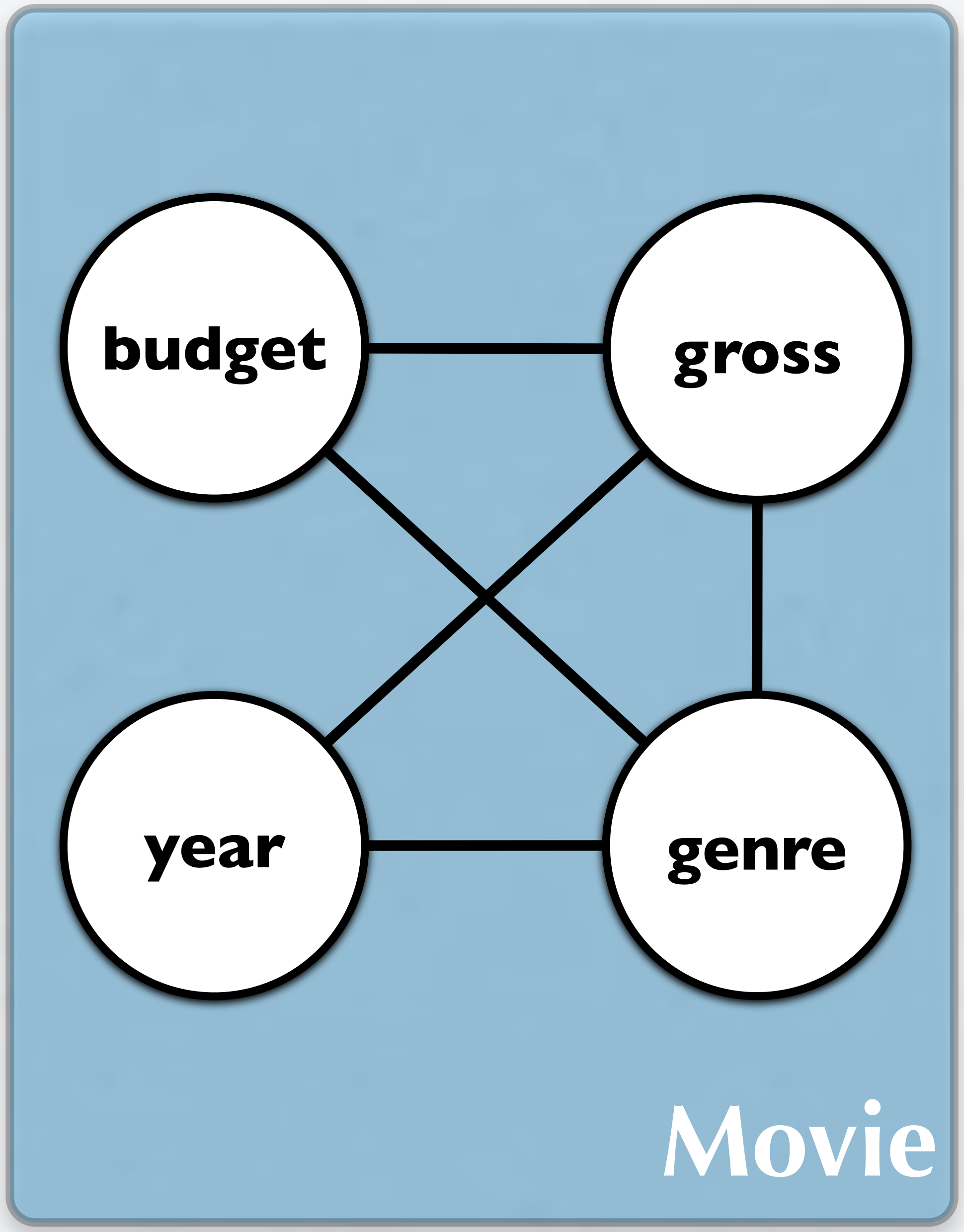
(Spirtes, Glymour, Scheines 1993; 2001)

The SGS algorithm

Phase I

Skeleton identification

Determine set of conditional independencies among all variables



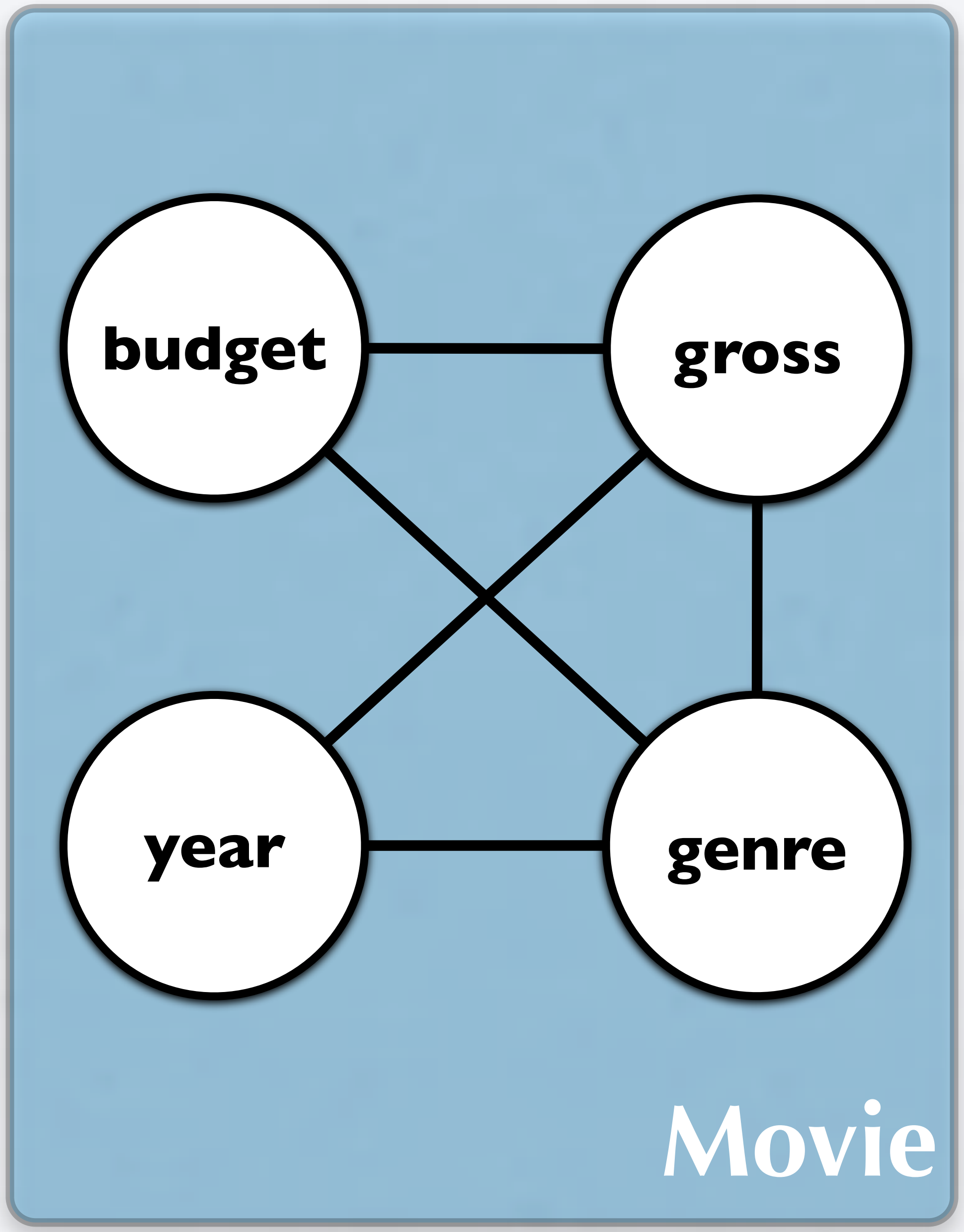
(Spirtes, Glymour, Scheines 1993; 2001)

The SGS algorithm

Phase I

Skeleton identification

Determine set of conditional independencies among all variables



$$\text{genre} \perp\!\!\!\perp \text{gross} \mid \text{budget} \ ?$$

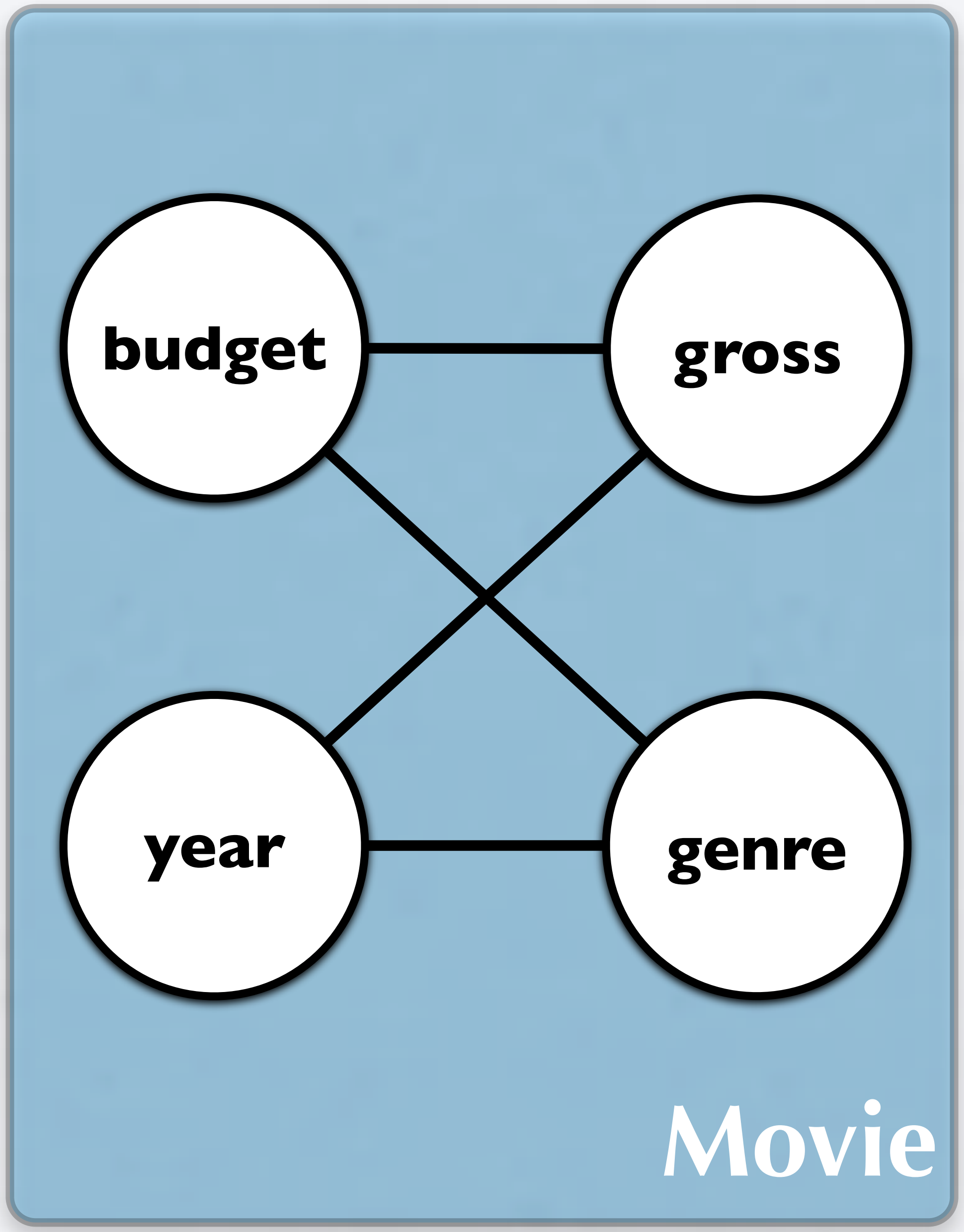
(Spirtes, Glymour, Scheines 1993; 2001)

The SGS algorithm

Phase I

Skeleton identification

Determine set of conditional independencies among all variables



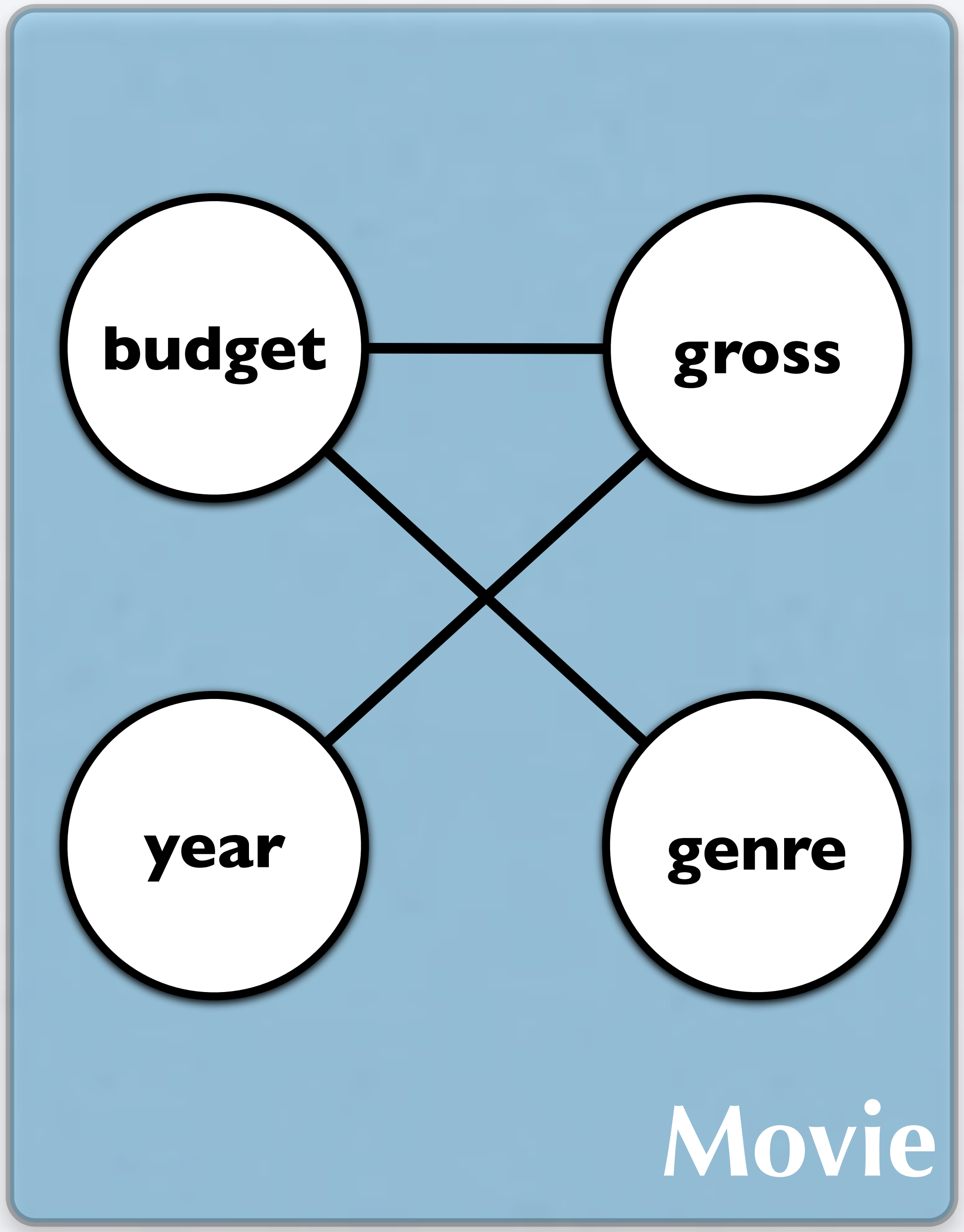
(Spirtes, Glymour, Scheines 1993; 2001)

The SGS algorithm

Phase I

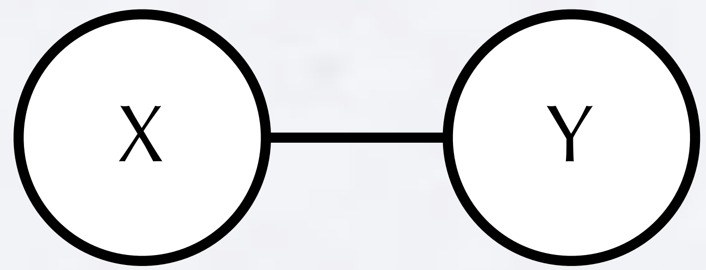
Skeleton identification

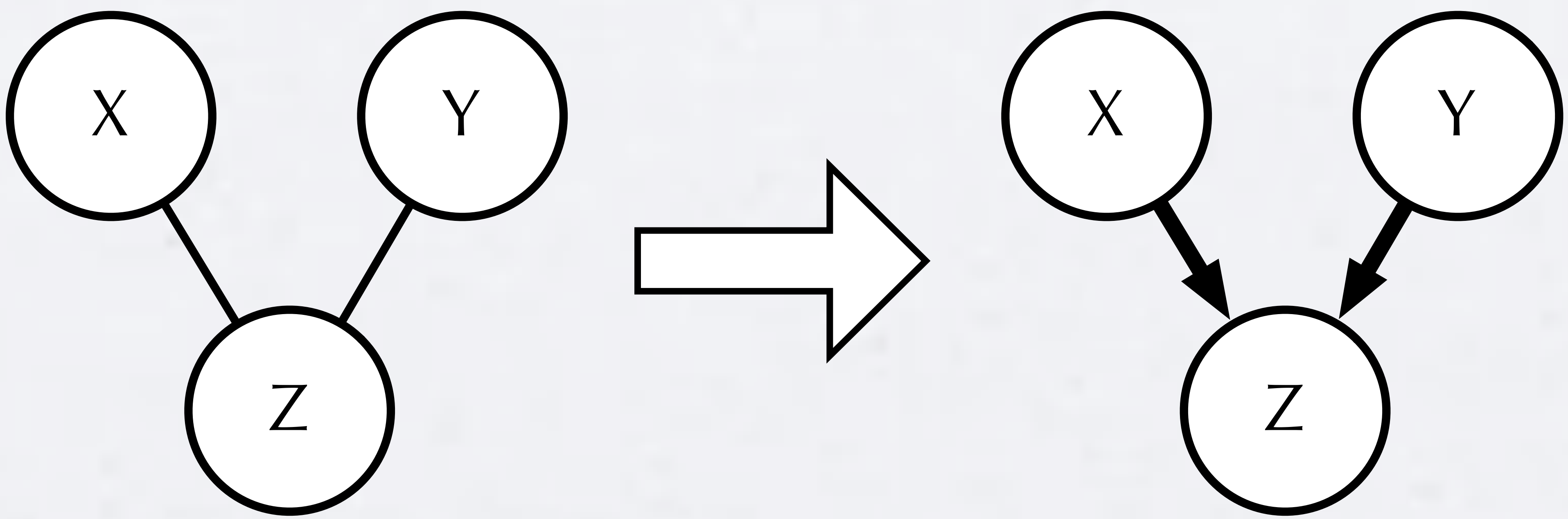
Determine set of conditional independencies among all variables



(Spirtes, Glymour, Scheines 1993; 2001)

Conditional independence





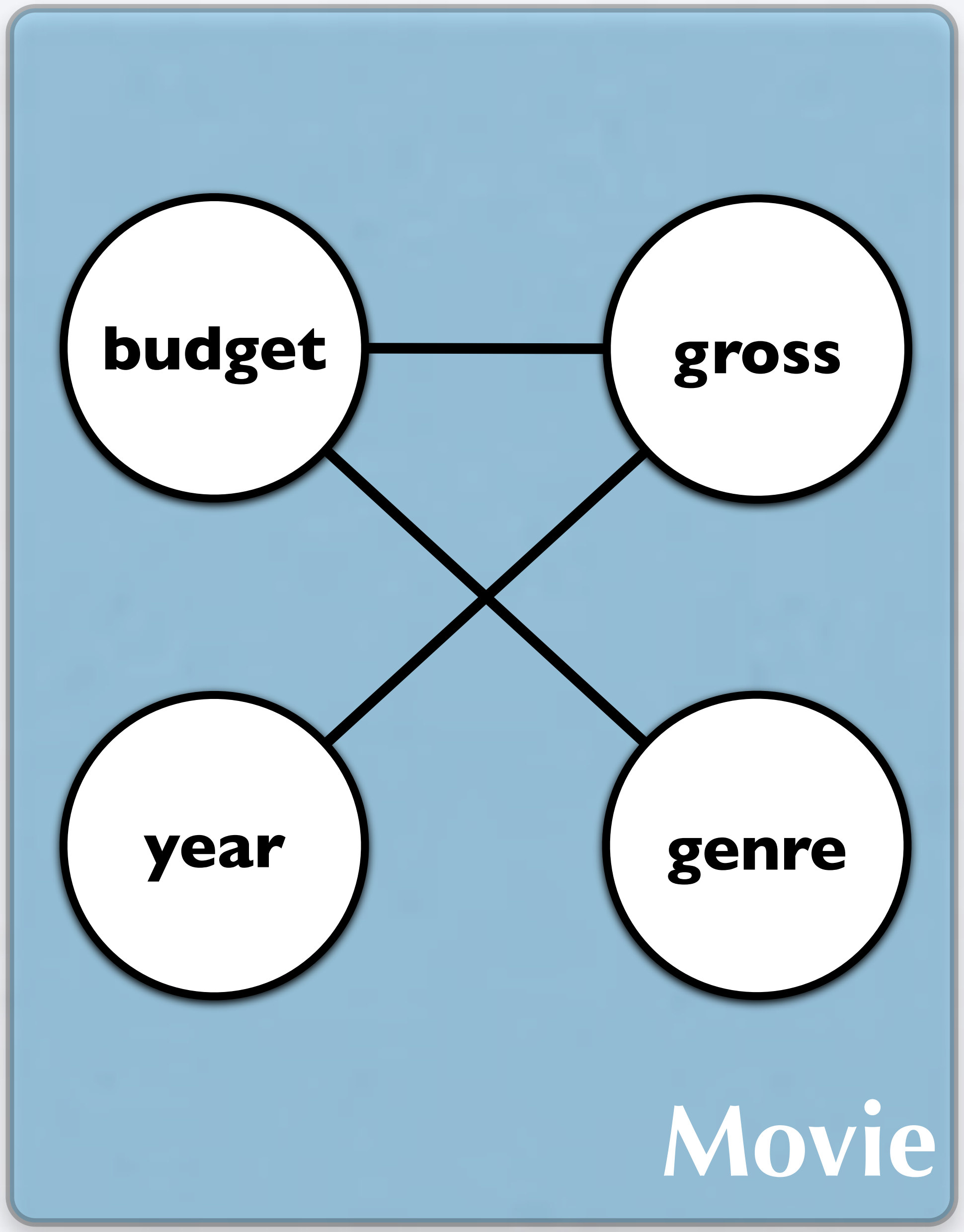
$$X \perp\!\!\!\perp Y \mid \mathbf{W},\ Z \notin \mathbf{W}$$

The SGS algorithm

Phase II

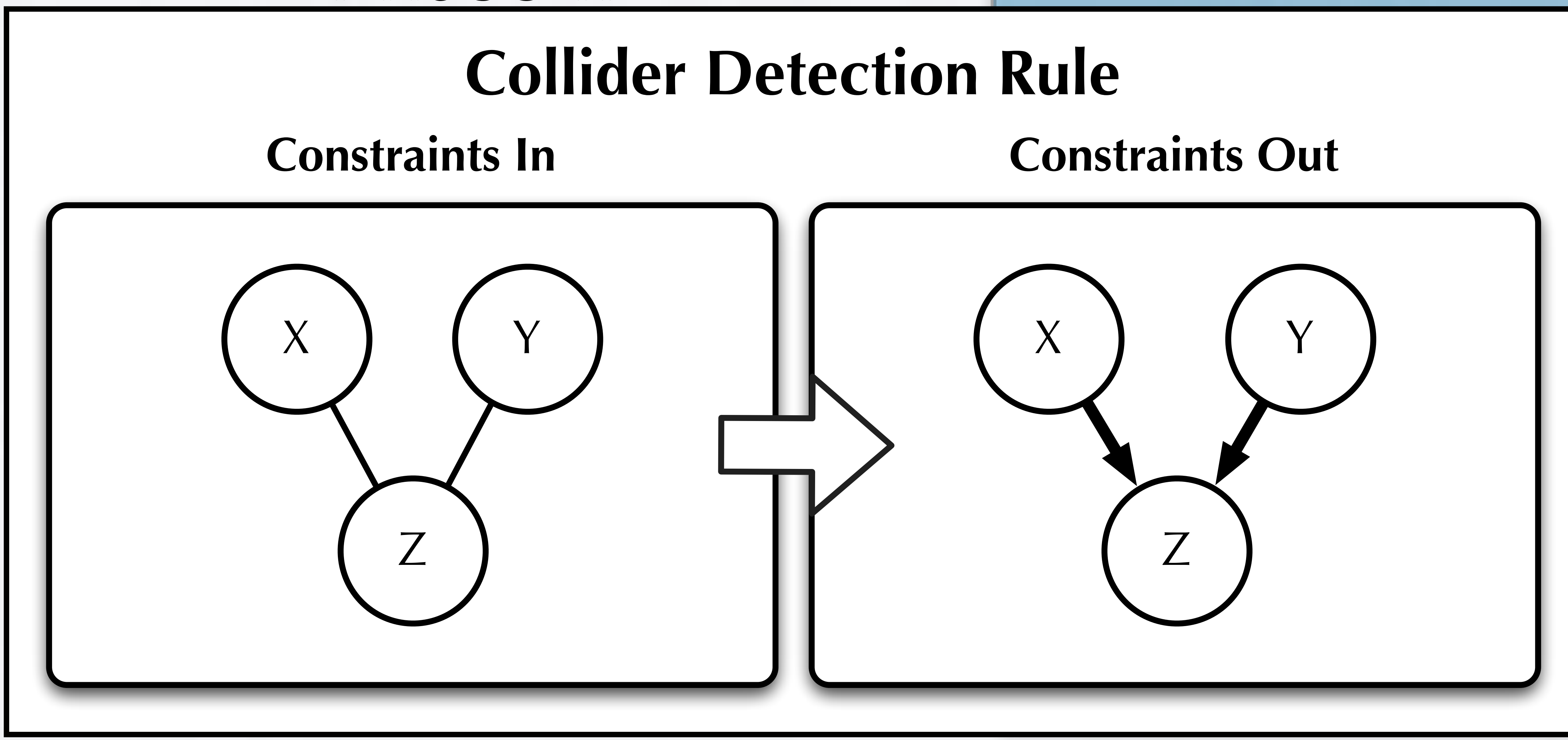
Edge orientation

Apply rules to uniquely determine causal structure consistent with patterns of association from Phase I



The SGS algorithm

Phase II

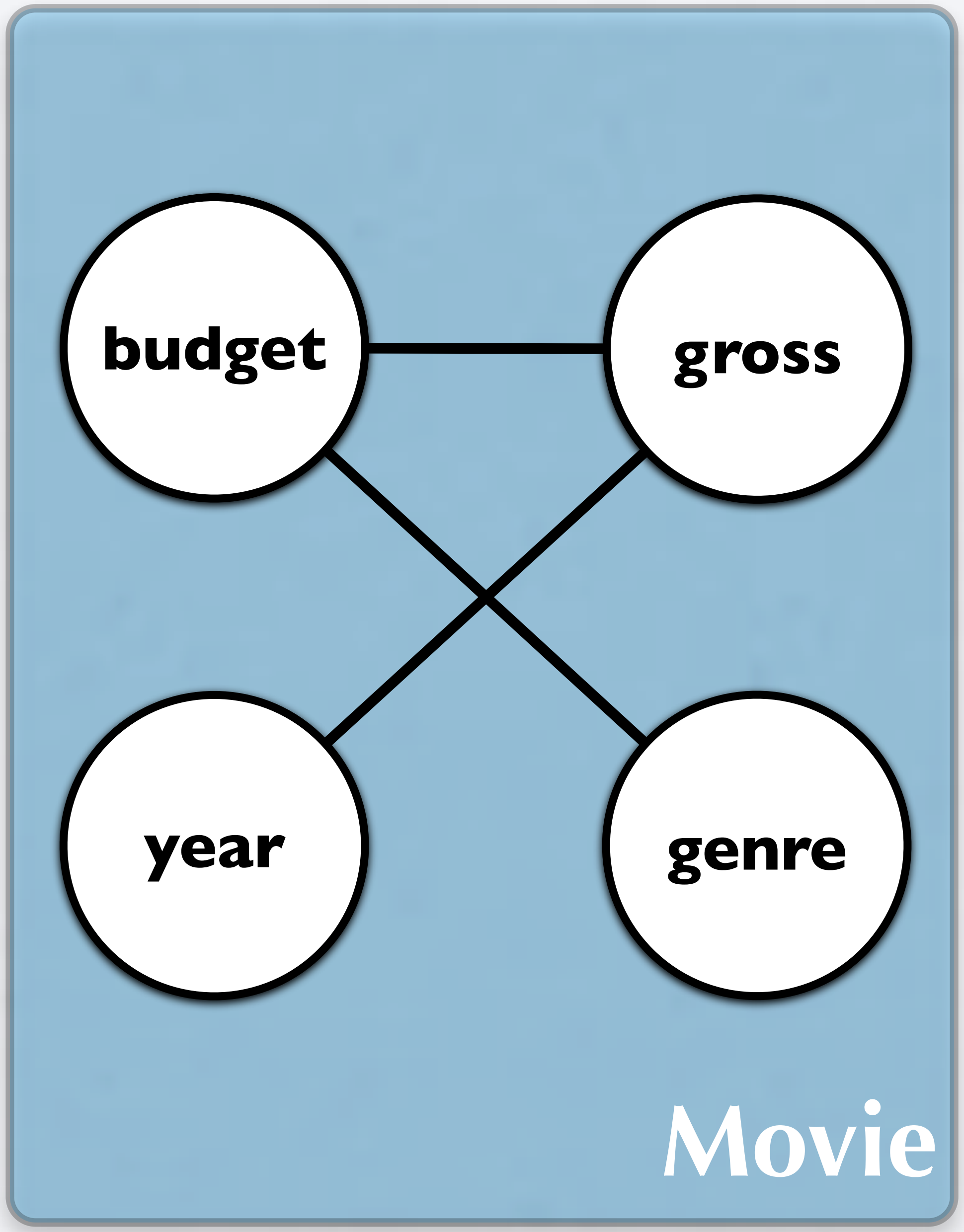


The SGS algorithm

Phase II

Edge orientation

Apply rules to uniquely determine causal structure consistent with patterns of association from Phase I



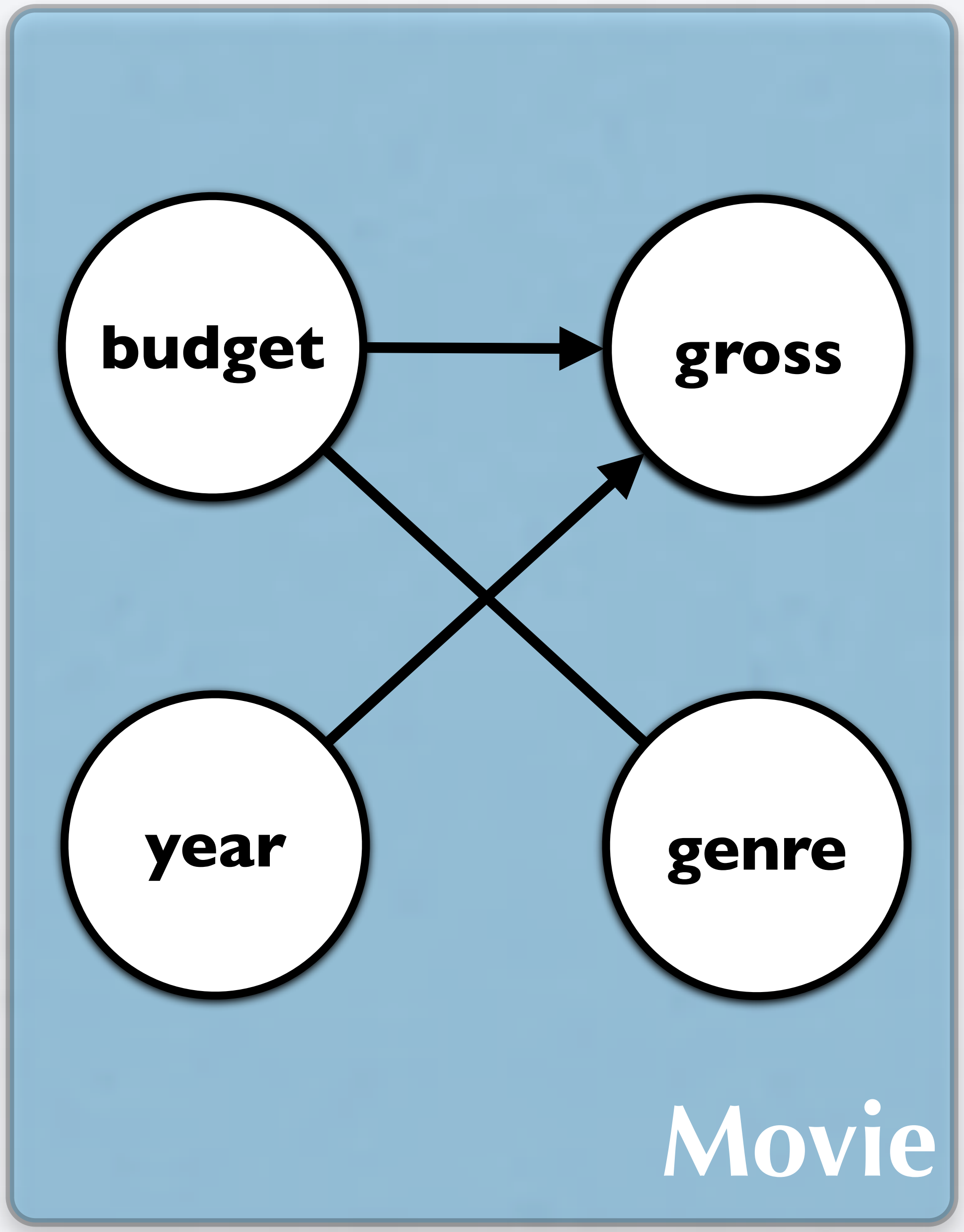
$$\text{year} \perp\!\!\!\perp \text{budget} \mid \{W\}$$
$$\text{gross} \notin W$$

The SGS algorithm

Phase II

Edge orientation

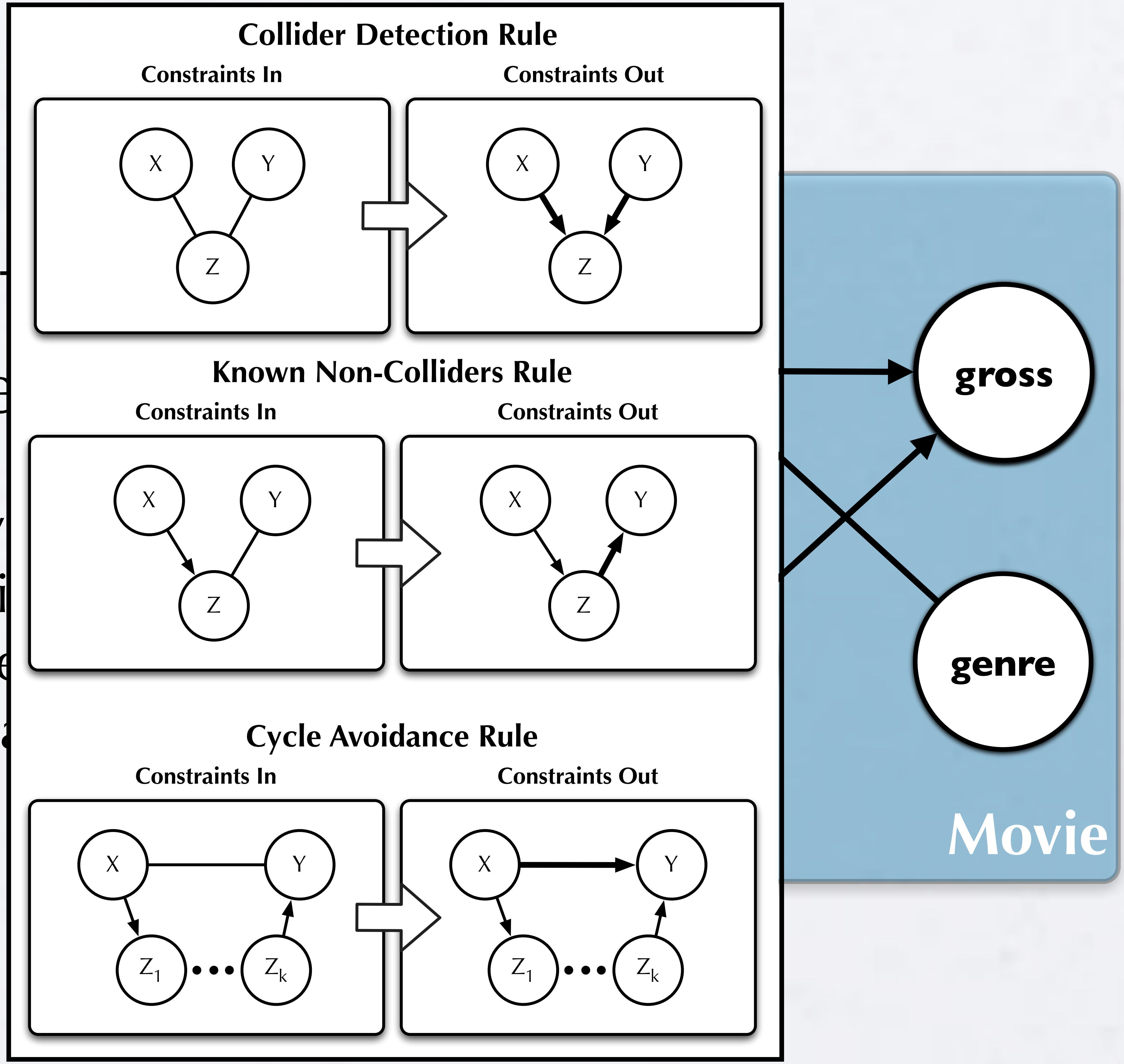
Apply rules to uniquely determine causal structure consistent with patterns of association from Phase I



$$\text{year} \perp\!\!\!\perp \text{budget} \mid \{W\}$$
$$\text{gross} \notin W$$

Edge

Apply
determi
consiste
associ

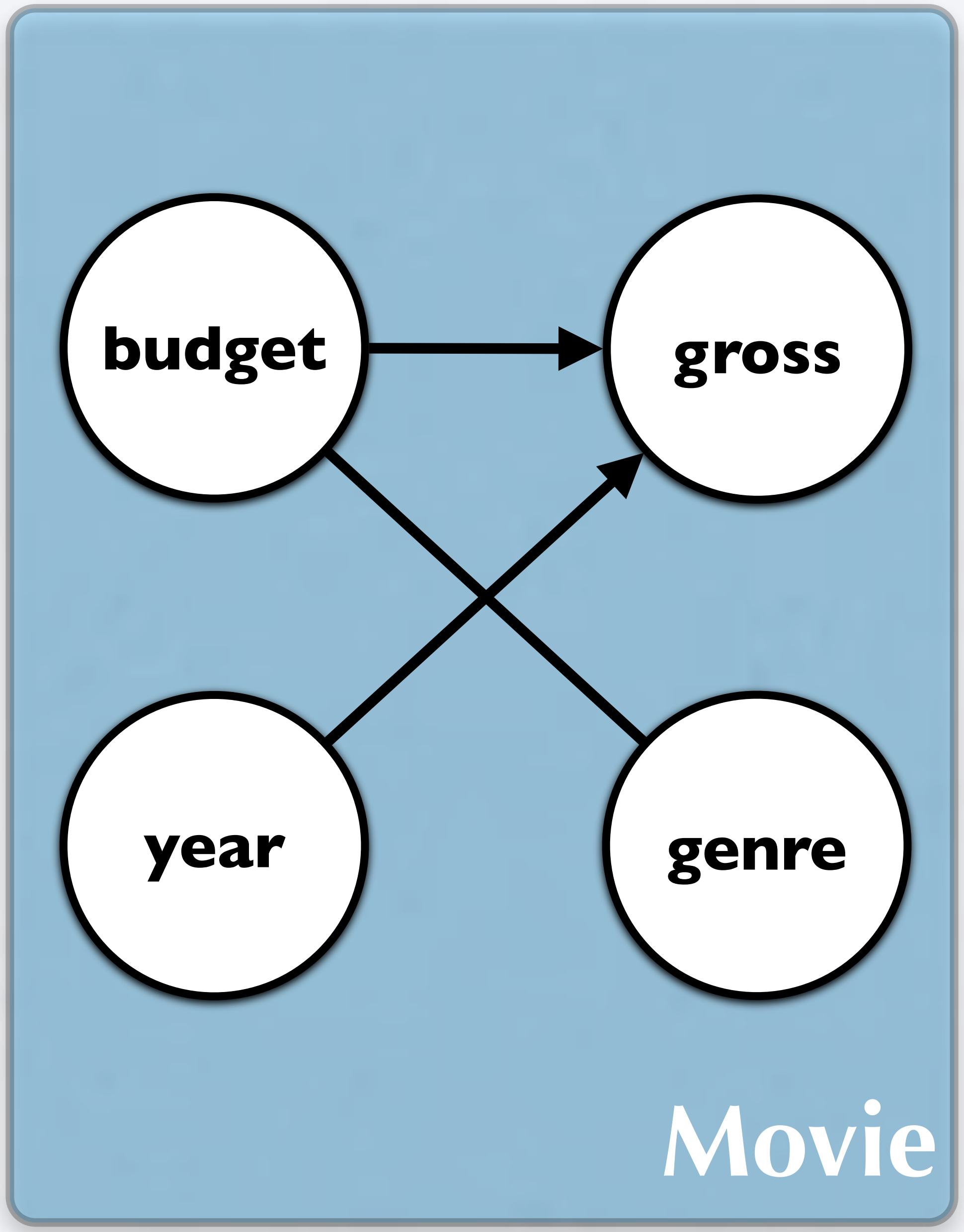


The SGS algorithm

Phase II

Edge orientation

Apply rules to uniquely determine causal structure consistent with patterns of association from Phase I



SGS correctly identifies a class of statistically indistinguishable causal models

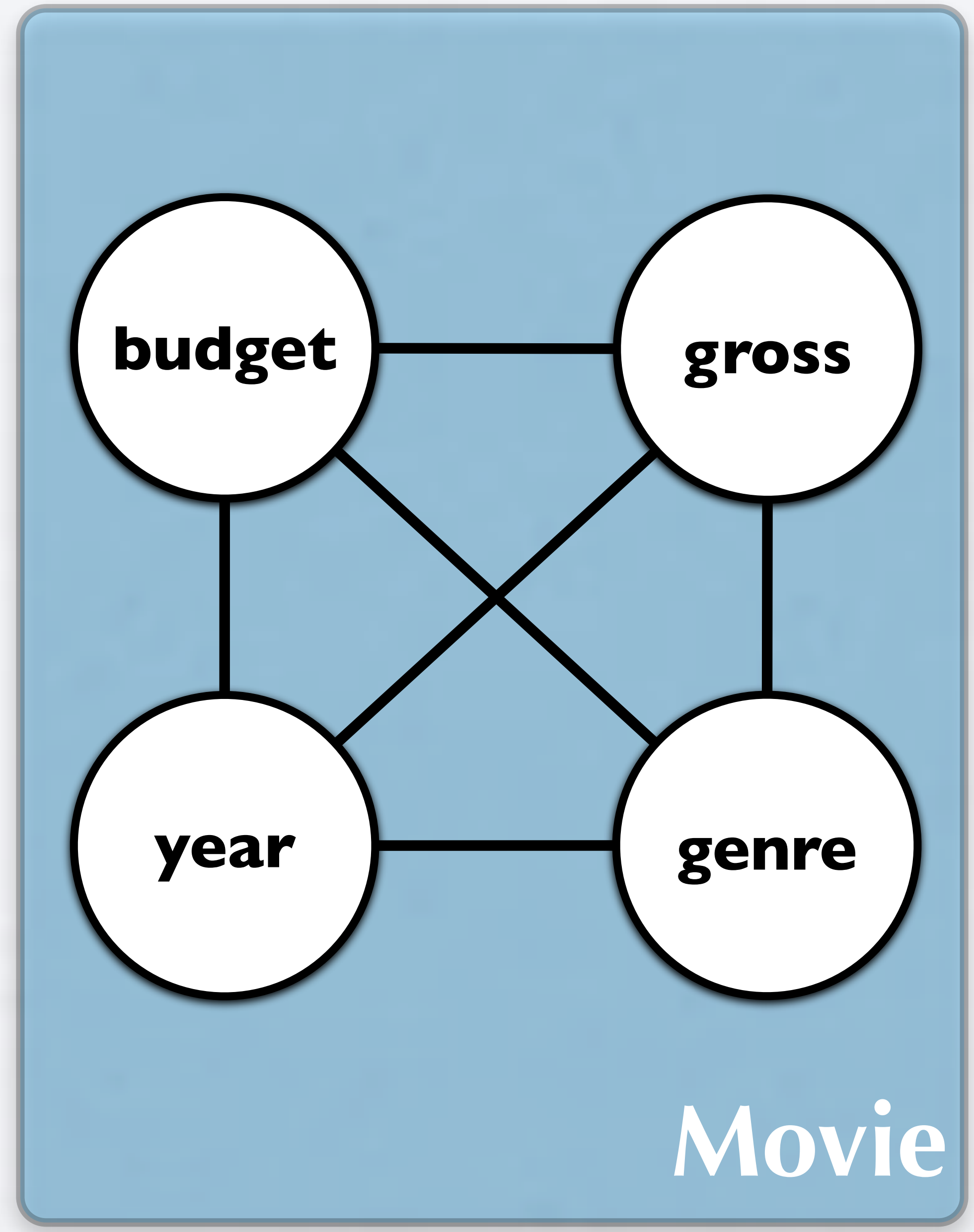
PC

The PC algorithm

Phase I

Skeleton identification

Determine set of conditional independencies among all variables



(Spirtes, Glymour, Scheines 1993; 2001)

The PC algorithm

$\mathrm{budget} \perp\!\!\!\perp \mathrm{genre} \mid \{\}$
$\mathrm{budget} \perp\!\!\!\perp \mathrm{gross} \mid \{\}$
$\mathrm{budget} \perp\!\!\!\perp \mathrm{year} \mid \{\}$
$\mathrm{genre} \perp\!\!\!\perp \mathrm{budget} \mid \{\}$
$\mathrm{genre} \perp\!\!\!\perp \mathrm{gross} \mid \{\}$
$\mathrm{genre} \perp\!\!\!\perp \mathrm{year} \mid \{\}$
$\mathrm{gross} \perp\!\!\!\perp \mathrm{budget} \mid \{\}$
$\mathrm{gross} \perp\!\!\!\perp \mathrm{genre} \mid \{\}$
$\mathrm{gross} \perp\!\!\!\perp \mathrm{year} \mid \{\}$
$\mathrm{year} \perp\!\!\!\perp \mathrm{budget} \mid \{\}$
$\mathrm{year} \perp\!\!\!\perp \mathrm{genre} \mid \{\}$
$\mathrm{year} \perp\!\!\!\perp \mathrm{gross} \mid \{\}$
$\mathrm{budget} \perp\!\!\!\perp \mathrm{genre} \mid \{\mathrm{gross}\}$
$\mathrm{budget} \perp\!\!\!\perp \mathrm{genre} \mid \{\mathrm{year}\}$
$\mathrm{budget} \perp\!\!\!\perp \mathrm{gross} \mid \{\mathrm{genre}\}$
$\mathrm{budget} \perp\!\!\!\perp \mathrm{gross} \mid \{\mathrm{year}\}$
$\mathrm{budget} \perp\!\!\!\perp \mathrm{year} \mid \{\mathrm{genre}\}$
$\mathrm{budget} \perp\!\!\!\perp \mathrm{year} \mid \{\mathrm{gross}\}$
$\mathrm{genre} \perp\!\!\!\perp \mathrm{budget} \mid \{\mathrm{gross}\}$
$\mathrm{genre} \perp\!\!\!\perp \mathrm{budget} \mid \{\mathrm{year}\}$
$\mathrm{genre} \perp\!\!\!\perp \mathrm{gross} \mid \{\mathrm{budget}\}$
$\mathrm{genre} \perp\!\!\!\perp \mathrm{gross} \mid \{\mathrm{year}\}$
$\mathrm{genre} \perp\!\!\!\perp \mathrm{year} \mid \{\mathrm{budget}\}$
$\mathrm{genre} \perp\!\!\!\perp \mathrm{year} \mid \{\mathrm{gross}\}$
$\mathrm{gross} \perp\!\!\!\perp \mathrm{budget} \mid \{\mathrm{genre}\}$
$\mathrm{gross} \perp\!\!\!\perp \mathrm{budget} \mid \{\mathrm{year}\}$
$\mathrm{gross} \perp\!\!\!\perp \mathrm{genre} \mid \{\mathrm{budget}\}$
$\mathrm{gross} \perp\!\!\!\perp \mathrm{genre} \mid \{\mathrm{year}\}$
$\mathrm{gross} \perp\!\!\!\perp \mathrm{year} \mid \{\mathrm{budget}\}$
$\mathrm{gross} \perp\!\!\!\perp \mathrm{year} \mid \{\mathrm{genre}\}$
$\mathrm{year} \perp\!\!\!\perp \mathrm{budget} \mid \{\mathrm{genre}\}$
$\mathrm{year} \perp\!\!\!\perp \mathrm{budget} \mid \{\mathrm{gross}\}$
$\mathrm{year} \perp\!\!\!\perp \mathrm{genre} \mid \{\mathrm{budget}\}$
$\mathrm{year} \perp\!\!\!\perp \mathrm{genre} \mid \{\mathrm{gross}\}$
$\mathrm{year} \perp\!\!\!\perp \mathrm{gross} \mid \{\mathrm{budget}\}$
$\mathrm{year} \perp\!\!\!\perp \mathrm{gross} \mid \{\mathrm{genre}\}$
$\mathrm{budget} \perp\!\!\!\perp \mathrm{genre} \mid \{\mathrm{gross}, \mathrm{year}\}$
$\mathrm{budget} \perp\!\!\!\perp \mathrm{gross} \mid \{\mathrm{genre}, \mathrm{year}\}$
$\mathrm{budget} \perp\!\!\!\perp \mathrm{year} \mid \{\mathrm{genre}, \mathrm{gross}\}$
$\mathrm{genre} \perp\!\!\!\perp \mathrm{budget} \mid \{\mathrm{gross}, \mathrm{year}\}$
$\mathrm{genre} \perp\!\!\!\perp \mathrm{gross} \mid \{\mathrm{budget}, \mathrm{year}\}$
$\mathrm{genre} \perp\!\!\!\perp \mathrm{year} \mid \{\mathrm{budget}, \mathrm{gross}\}$
$\mathrm{gross} \perp\!\!\!\perp \mathrm{budget} \mid \{\mathrm{genre}, \mathrm{year}\}$
$\mathrm{gross} \perp\!\!\!\perp \mathrm{genre} \mid \{\mathrm{budget}, \mathrm{year}\}$
$\mathrm{gross} \perp\!\!\!\perp \mathrm{year} \mid \{\mathrm{budget}, \mathrm{genre}\}$
$\mathrm{year} \perp\!\!\!\perp \mathrm{budget} \mid \{\mathrm{genre}, \mathrm{gross}\}$
$\mathrm{year} \perp\!\!\!\perp \mathrm{genre} \mid \{\mathrm{budget}, \mathrm{gross}\}$
$\mathrm{year} \perp\!\!\!\perp \mathrm{gross} \mid \{\mathrm{budget}, \mathrm{genre}\}$

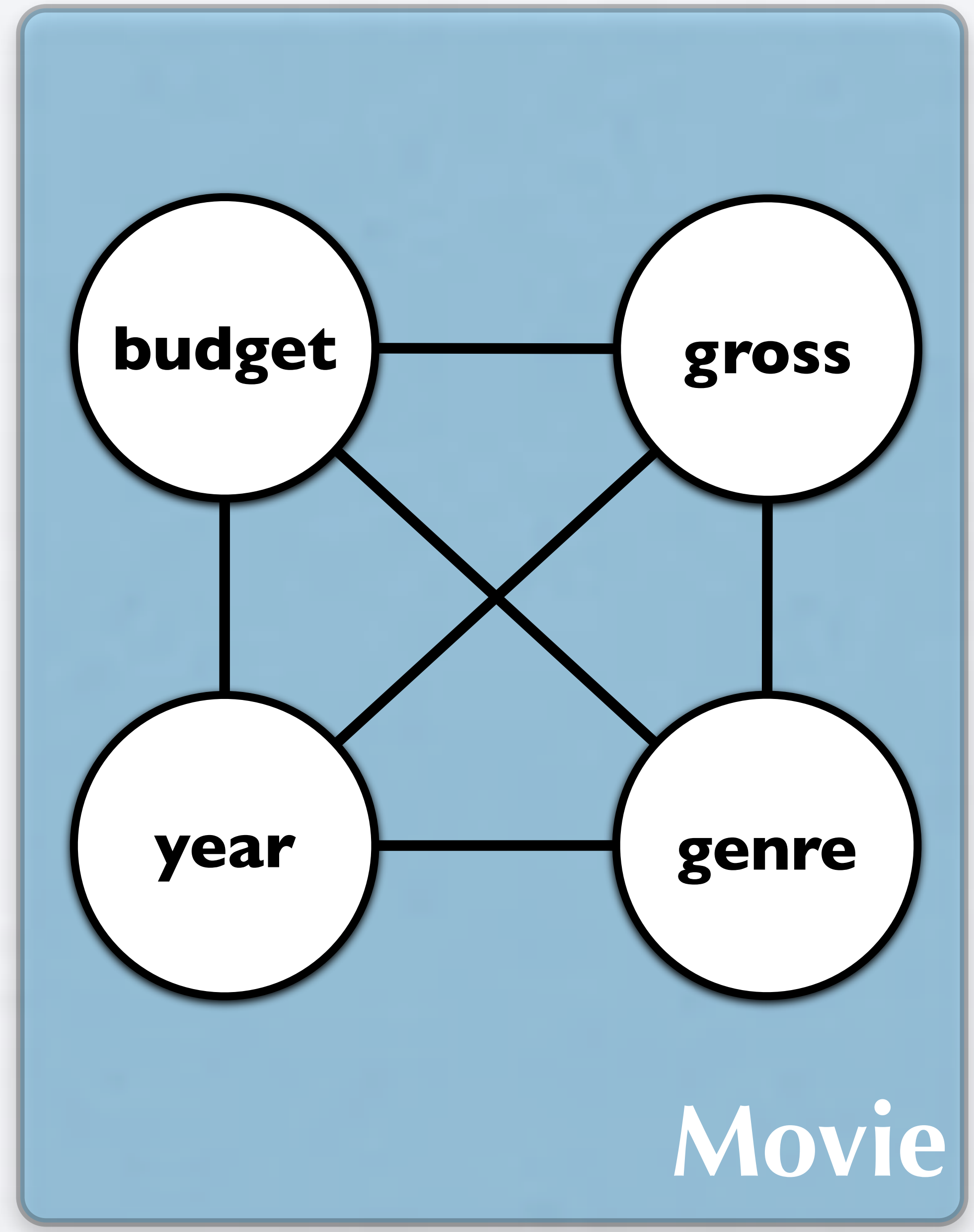
(Spirtes, Glymour, Scheines 1993; 2001)

The PC algorithm

Phase I

Skeleton identification

Determine set of conditional independencies among all variables



$$\text{budget} \perp\!\!\!\perp \text{year} \ ?$$

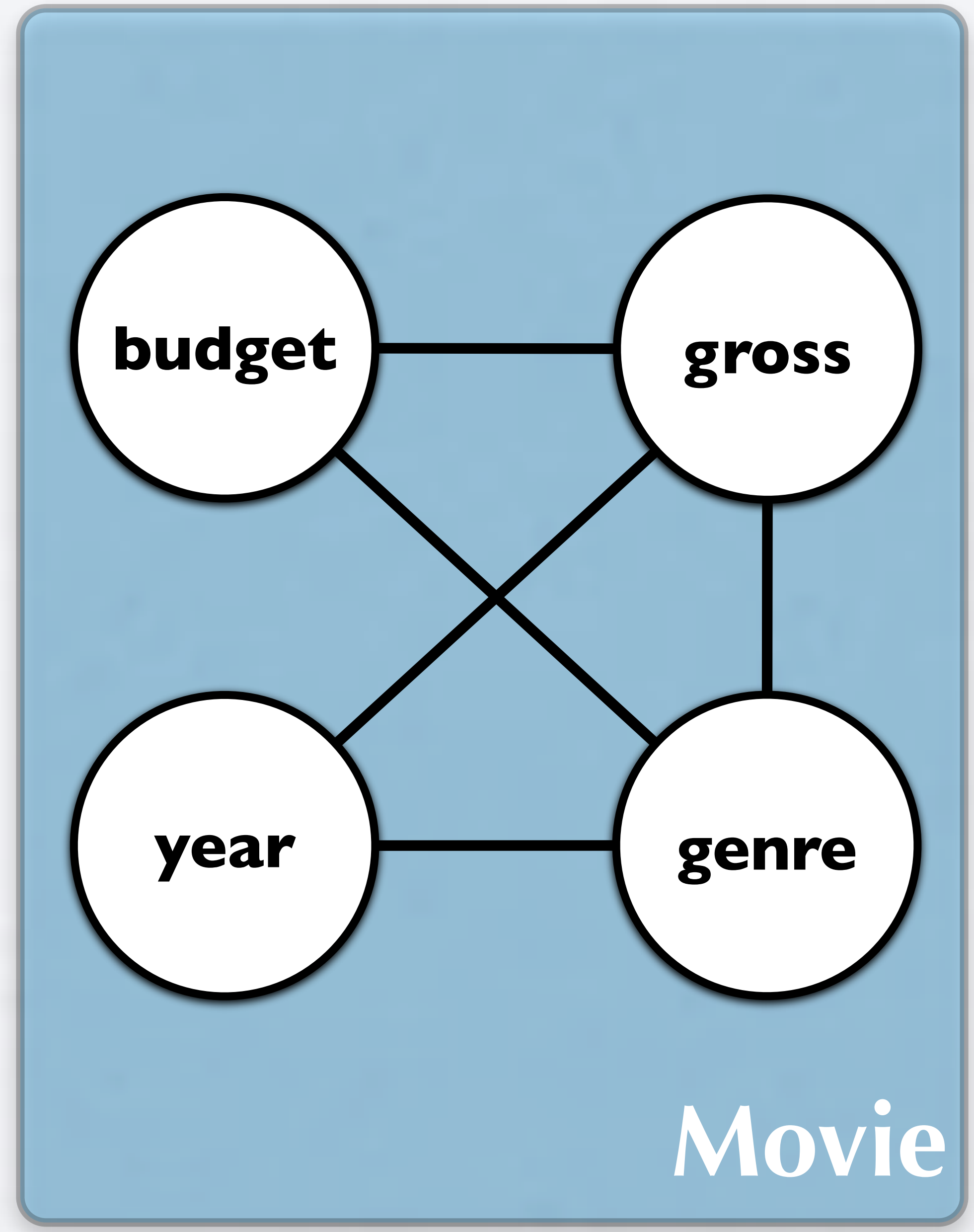
(Spirtes, Glymour, Scheines 1993; 2001)

The PC algorithm

Phase I

Skeleton identification

Determine set of conditional independencies among all variables



$$\text{budget} \perp\!\!\!\perp \text{year} \ ?$$

(Spirtes, Glymour, Scheines 1993; 2001)

The PC algorithm

budget $\perp\!\!\!\perp$ genre | {}
budget $\perp\!\!\!\perp$ gross | {}
~~budget $\perp\!\!\!\perp$ year | {}~~
genre $\perp\!\!\!\perp$ budget | {}
genre $\perp\!\!\!\perp$ gross | {}
genre $\perp\!\!\!\perp$ year | {}
gross $\perp\!\!\!\perp$ budget | {}
gross $\perp\!\!\!\perp$ genre | {}
gross $\perp\!\!\!\perp$ year | {}
~~year $\perp\!\!\!\perp$ budget | {}~~
year $\perp\!\!\!\perp$ genre | {}
year $\perp\!\!\!\perp$ gross | {}
budget $\perp\!\!\!\perp$ genre | {gross}
~~budget $\perp\!\!\!\perp$ genre | {year}~~
budget $\perp\!\!\!\perp$ gross | {genre}
~~budget $\perp\!\!\!\perp$ gross | {year}~~

~~budget $\perp\!\!\!\perp$ year | {genre}~~
~~budget $\perp\!\!\!\perp$ year | {gross}~~
genre $\perp\!\!\!\perp$ budget | {gross}
~~genre $\perp\!\!\!\perp$ budget | {year}~~
genre $\perp\!\!\!\perp$ gross | {budget}
genre $\perp\!\!\!\perp$ gross | {year}
genre $\perp\!\!\!\perp$ year | {budget}
genre $\perp\!\!\!\perp$ year | {gross}
gross $\perp\!\!\!\perp$ budget | {genre}
~~gross $\perp\!\!\!\perp$ budget | {year}~~
gross $\perp\!\!\!\perp$ genre | {budget}
gross $\perp\!\!\!\perp$ genre | {year}
~~gross $\perp\!\!\!\perp$ year | {budget}~~
gross $\perp\!\!\!\perp$ year | {genre}
~~year $\perp\!\!\!\perp$ budget | {genre}~~
~~year $\perp\!\!\!\perp$ budget | {gross}~~
~~year $\perp\!\!\!\perp$ genre | {budget}~~

year $\perp\!\!\!\perp$ genre | {gross}
~~year $\perp\!\!\!\perp$ gross | {budget}~~
year $\perp\!\!\!\perp$ gross | {genre}
~~budget $\perp\!\!\!\perp$ genre | {gross, year}~~
~~budget $\perp\!\!\!\perp$ gross | {genre, year}~~
~~budget $\perp\!\!\!\perp$ year | {genre, gross}~~
~~genre $\perp\!\!\!\perp$ budget | {gross, year}~~
~~genre $\perp\!\!\!\perp$ gross | {budget, year}~~
~~genre $\perp\!\!\!\perp$ year | {budget, gross}~~
~~gross $\perp\!\!\!\perp$ budget | {genre, year}~~
~~gross $\perp\!\!\!\perp$ genre | {budget, year}~~
~~gross $\perp\!\!\!\perp$ year | {budget, genre}~~
~~year $\perp\!\!\!\perp$ budget | {genre, gross}~~
~~year $\perp\!\!\!\perp$ genre | {budget, gross}~~
~~year $\perp\!\!\!\perp$ gross | {budget, genre}~~

budget $\perp\!\!\!\perp$ year ?

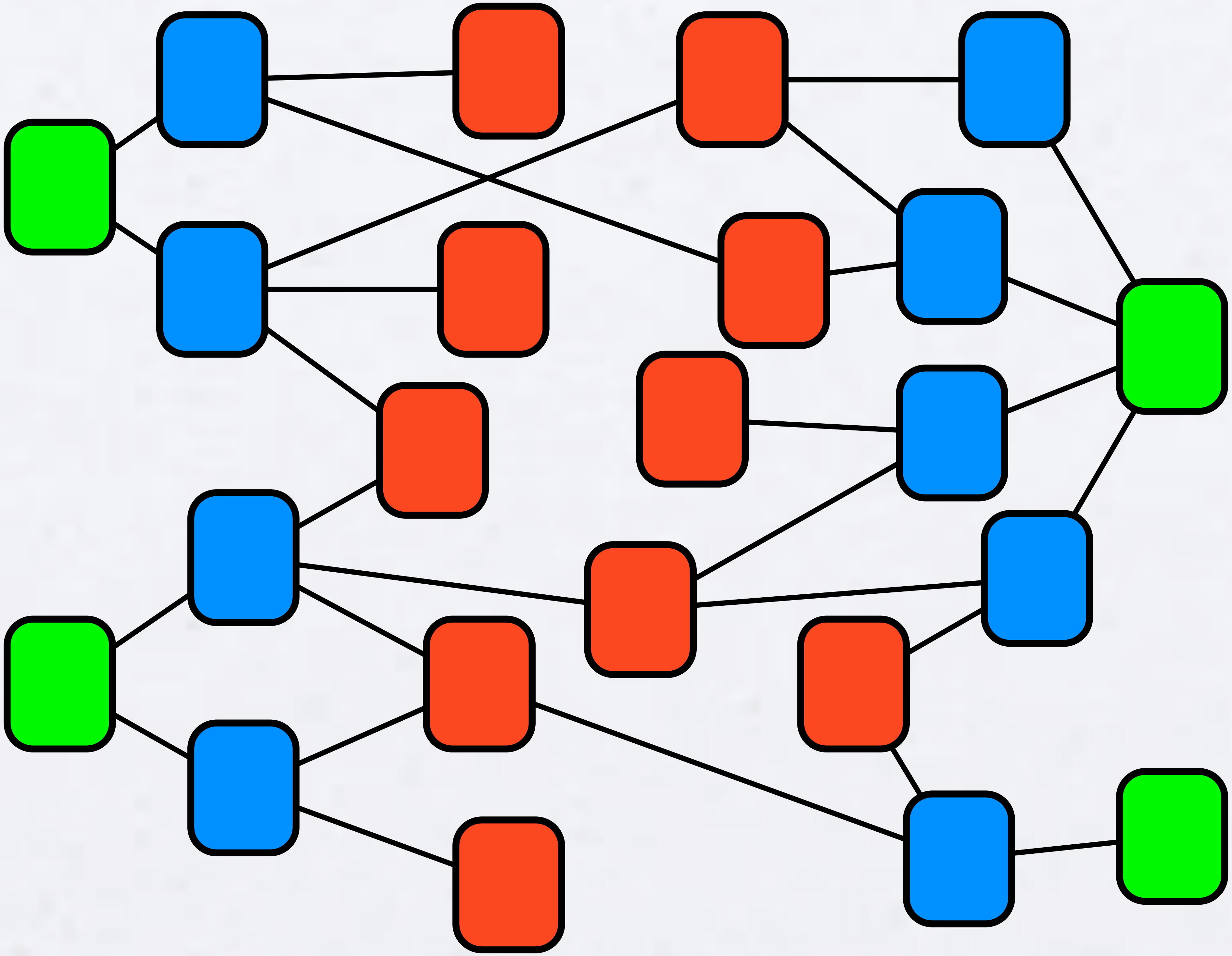
(Spirtes, Glymour, Scheines 1993; 2001)

Other propositional algorithms

- Relax/decompose faithfulness condition
  - Conservative PC (CPC) (Ramsey, Zhang, Spirtes 2006)
- Remove causal sufficiency assumption
  - Causal Inference (CI)
  - Fast Causal Inference (FCI) (Spirtes, Glymour, Scheines 1993; 2001)
- Practical modifications
  - Modified PC (Abellan, Gomez-Olmedo, Moral 2006)
  - POWER (Fast, Hay, Jensen 2008)
- Hybrid algorithms
  - MMHC (Tsamardinos, Brown, Aliferis 2006)
  - RELAX (Fast 2009)

RPC

Relational data



Relational database

movie id	movie	budget	gross	genre
1	Goodfellas	25M	47M	crime
2	My Cousin Vinny	11M	64M	comedy
	...	...	...	...

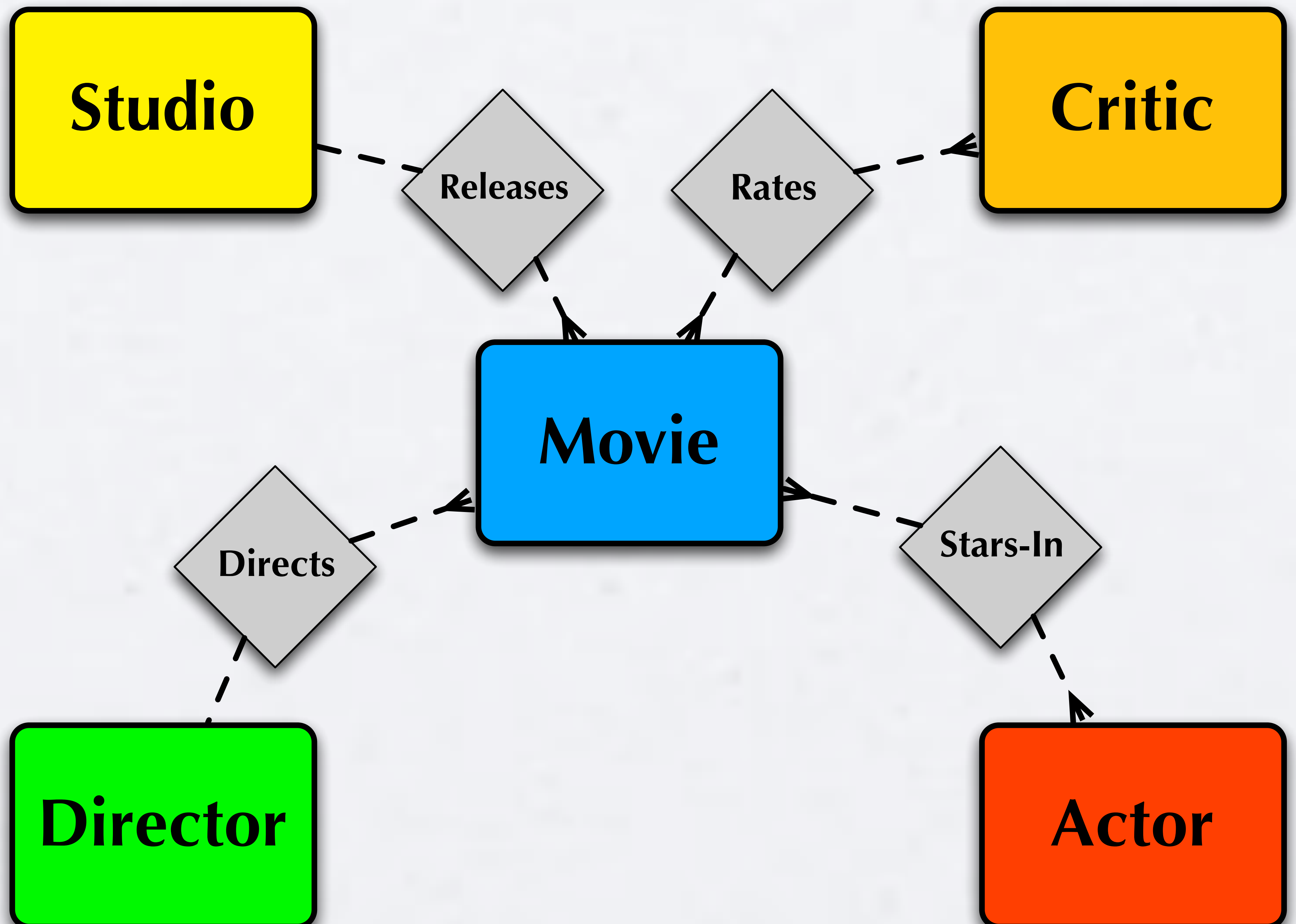
movie id	director id
1	1
2	2
...	...

director id	director
1	Martin Scorsese
2	Jonathan Lynn
...	...

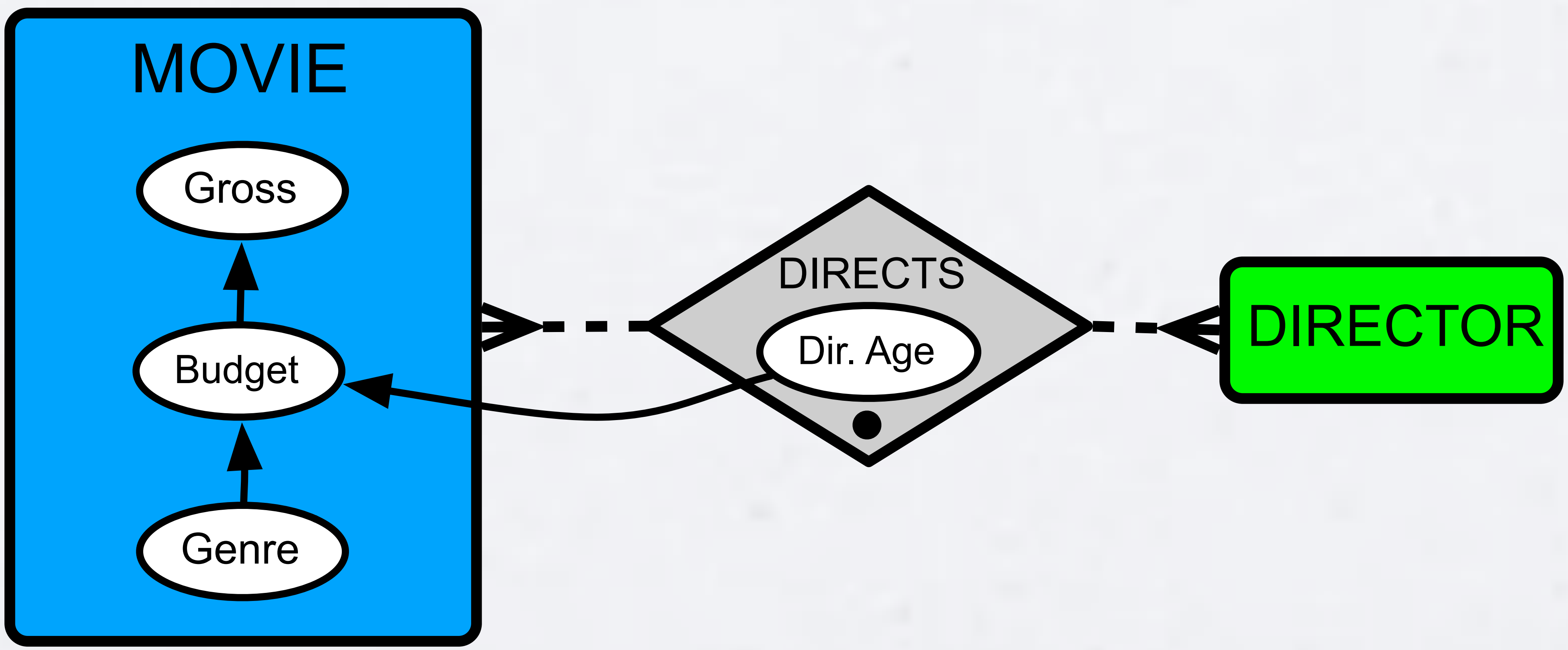
movie id	actor id
1	1
2	2
...	...

actor id	actor
1	Robert De Niro
2	Joe Pesci
...	...

Relational data representation

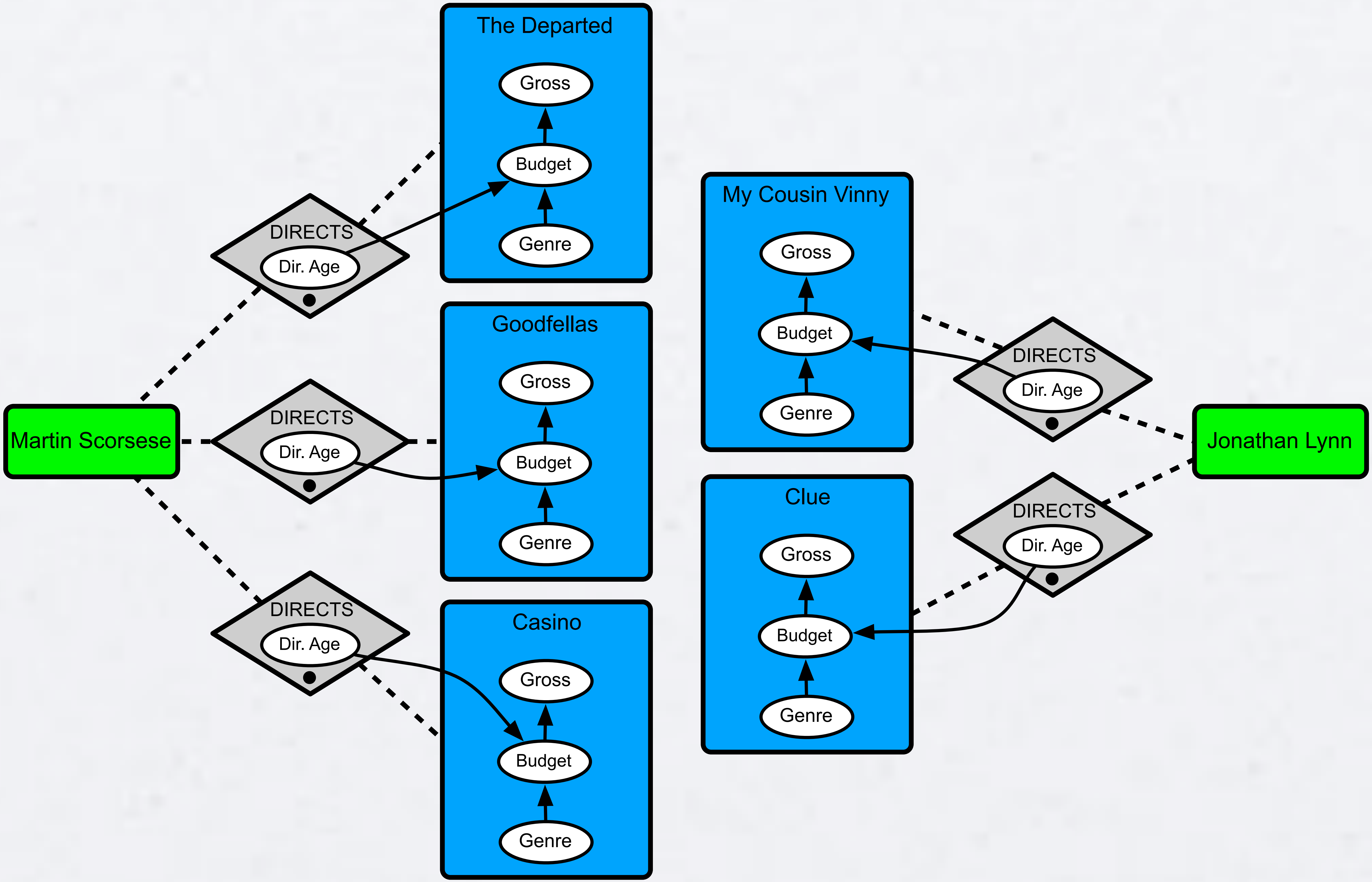


Directed Acyclic Probabilistic Entity-Relationship (DAPER) Model

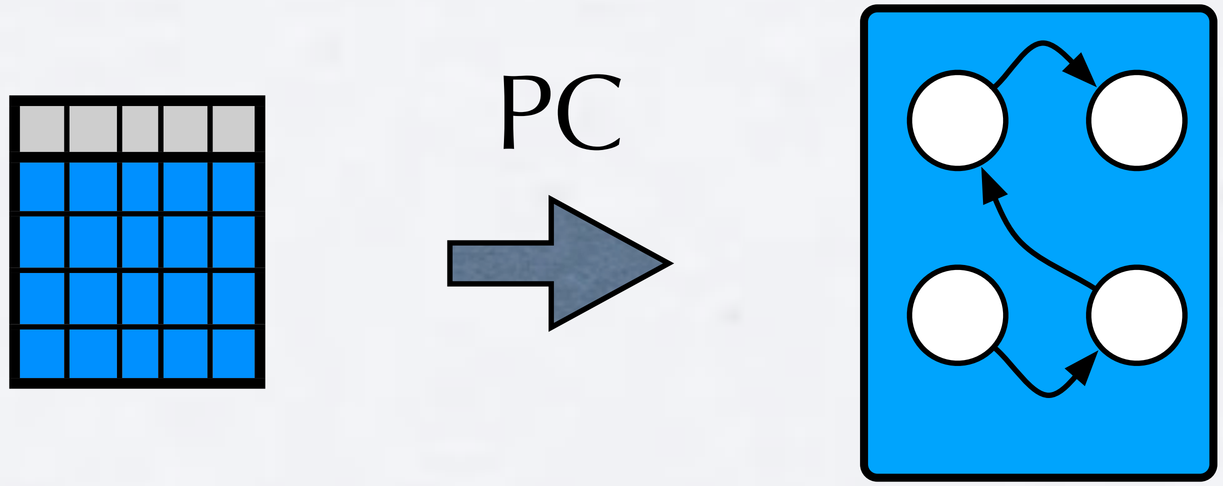


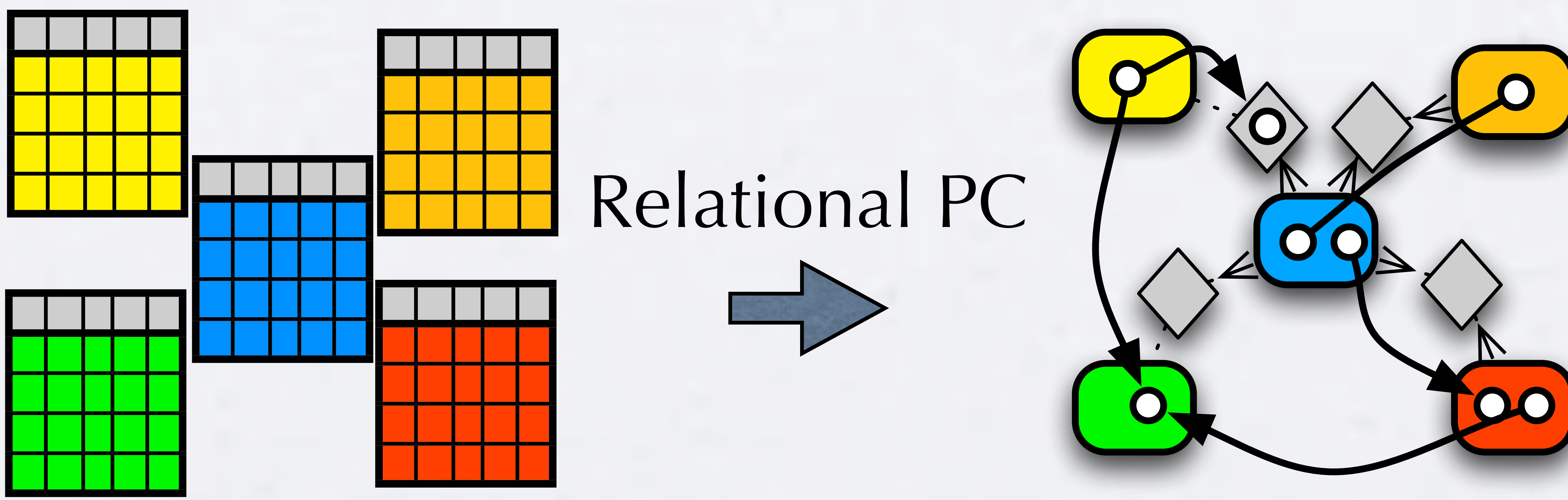
(Heckerman, Meek, Koller 2007)

DAPER ground graph



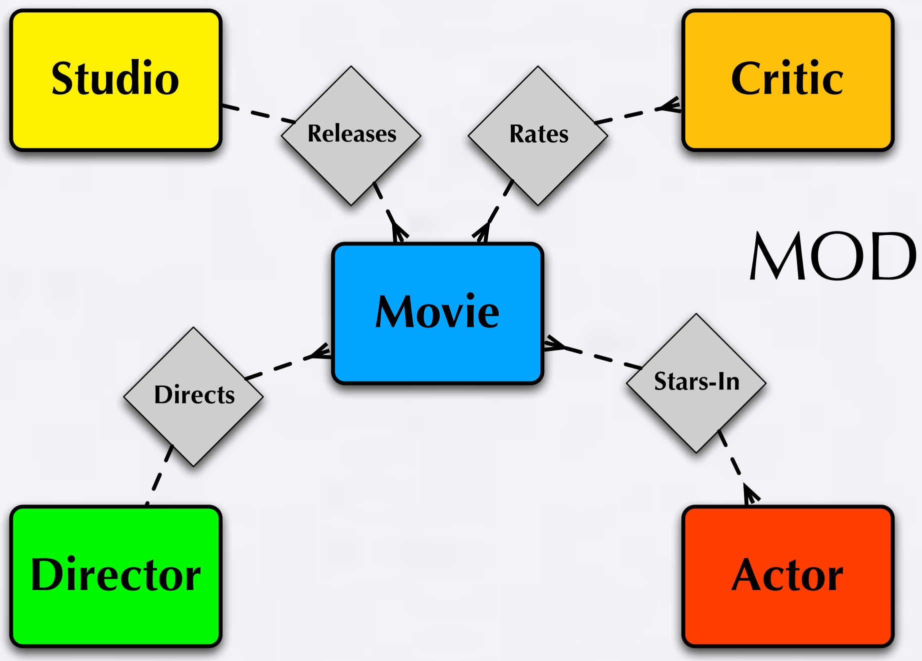
Relational extension of PC





(Maier, Taylor, Oktay, Jensen 2010)

Consequences of relational data



[Movie].budget

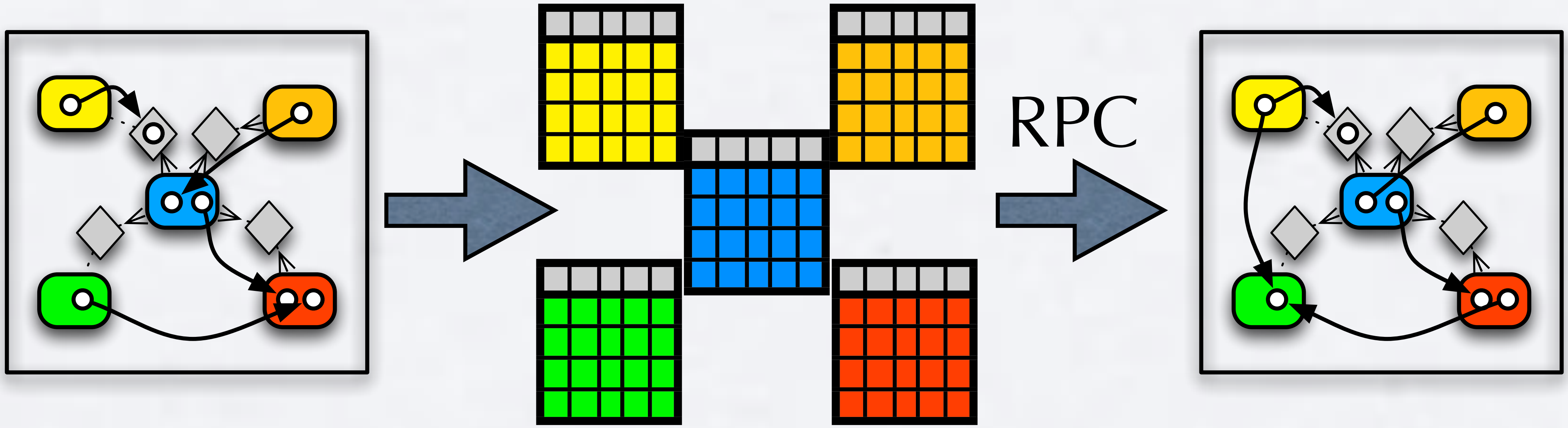
MODE([Movie Directs Director Directs Movie]).genre

COUNT([Movie Stars-In Actor])

EXISTS([Rates])

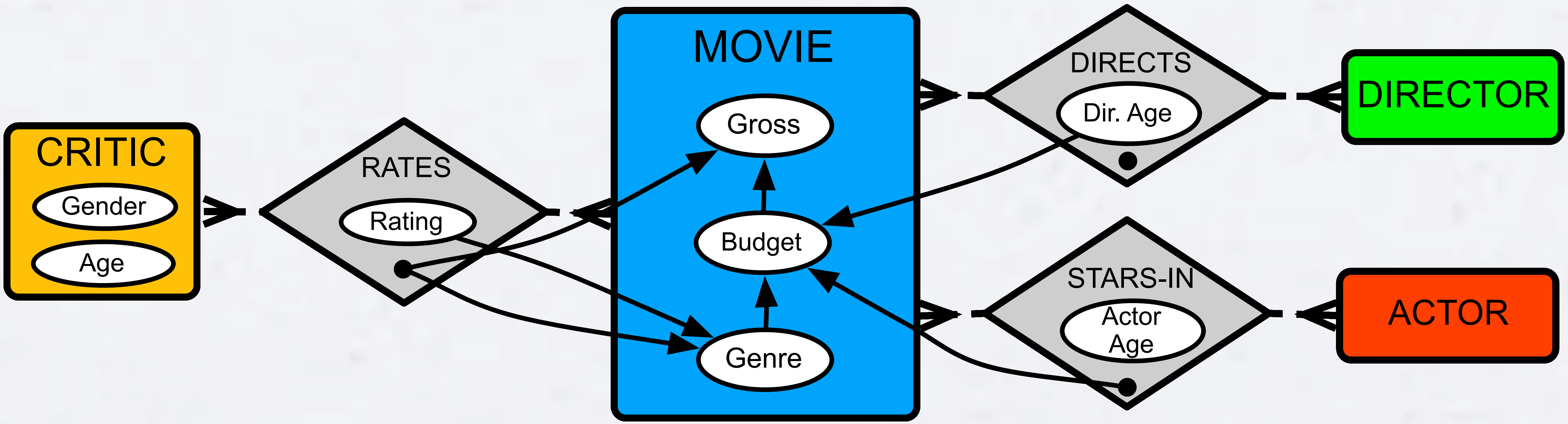
1. Increased **space of potential dependencies**
   - Variables from **multiple entities and relationships**
   - **Aggregates**
   - **Structural variables**
2. **New constraints** derived from relationship existence

Evaluating causal algorithms



1. Higher **power** yields more dependencies.
2. **Chain reactions** occur.
3. At best, identifies the class of **statistically indistinguishable** models.

Causal Model of MovieLens



Relational Blocking

Relational blocking defined

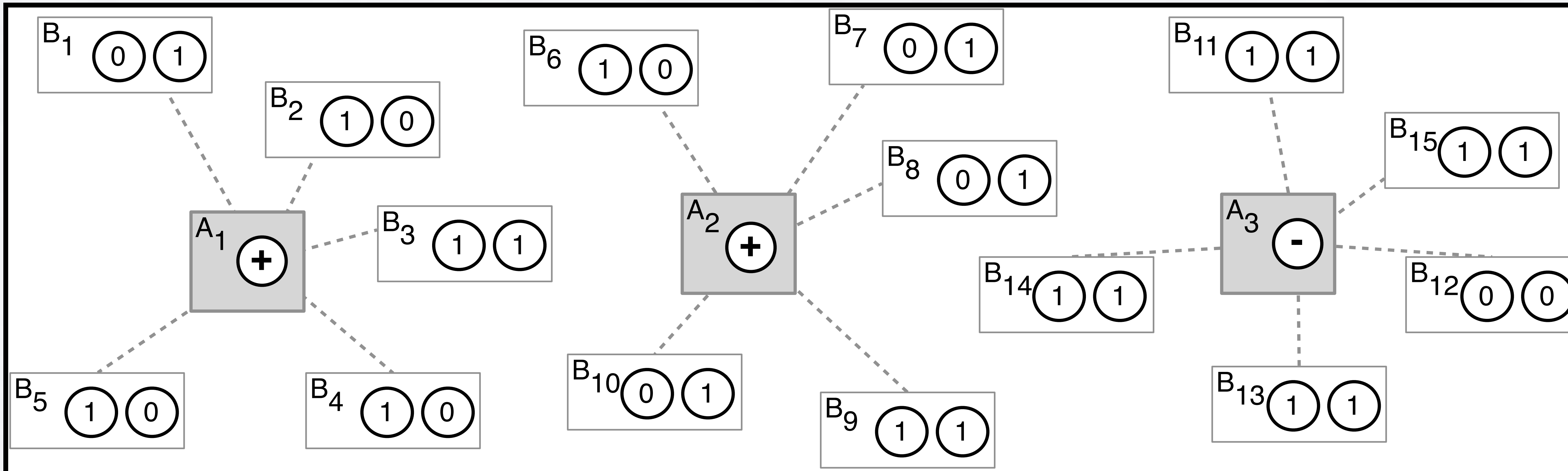
- Let *A* and *B* be two entity sets in a *k*-partite network
- A **block** contains a set of *B* entities linked to a common *A* entity
- Let *ID* be the unique identifier of a block, and let *X* and *Y* be two attributes of *B*
- **Relational blocking** is a process that evaluates

$$X \perp\!\!\!\perp Y \mid ID$$

by grouping *B* entities into disjoint blocks

(Rattigan, Maier, Jensen 2011)

Blocking vs Conditioning



Z = +

	X=0	1
Y=0	0	4
1	4	2

Z = -

	X=0	1
Y=0	1	0
1	0	4

Conditioning

A_1

	X=0	1
Y=0	0	3
1	1	1

A_2

	X=0	1
Y=0	0	1
1	3	1

A_3

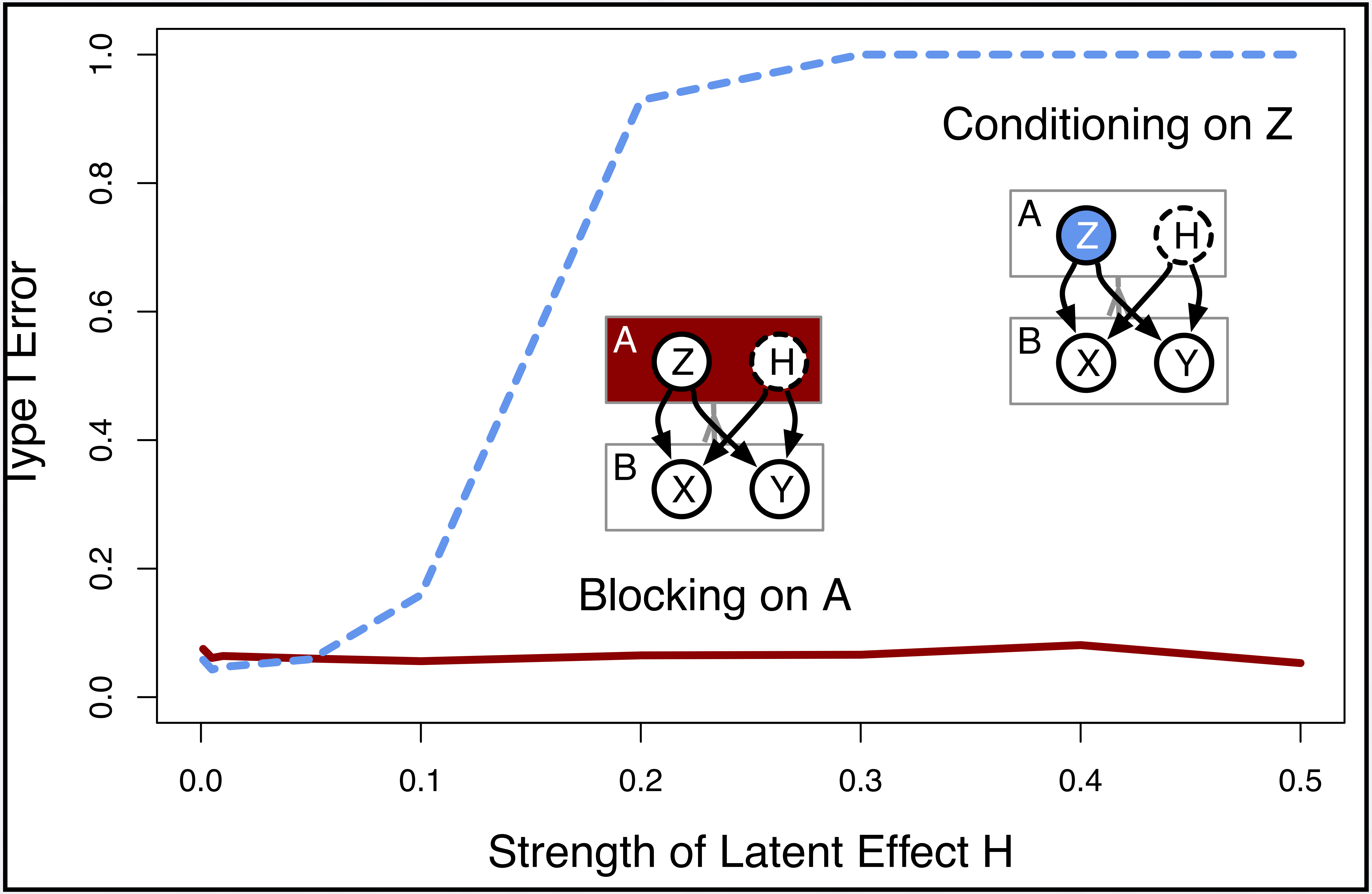
	X=0	1
Y=0	1	0
1	0	4

Blocking

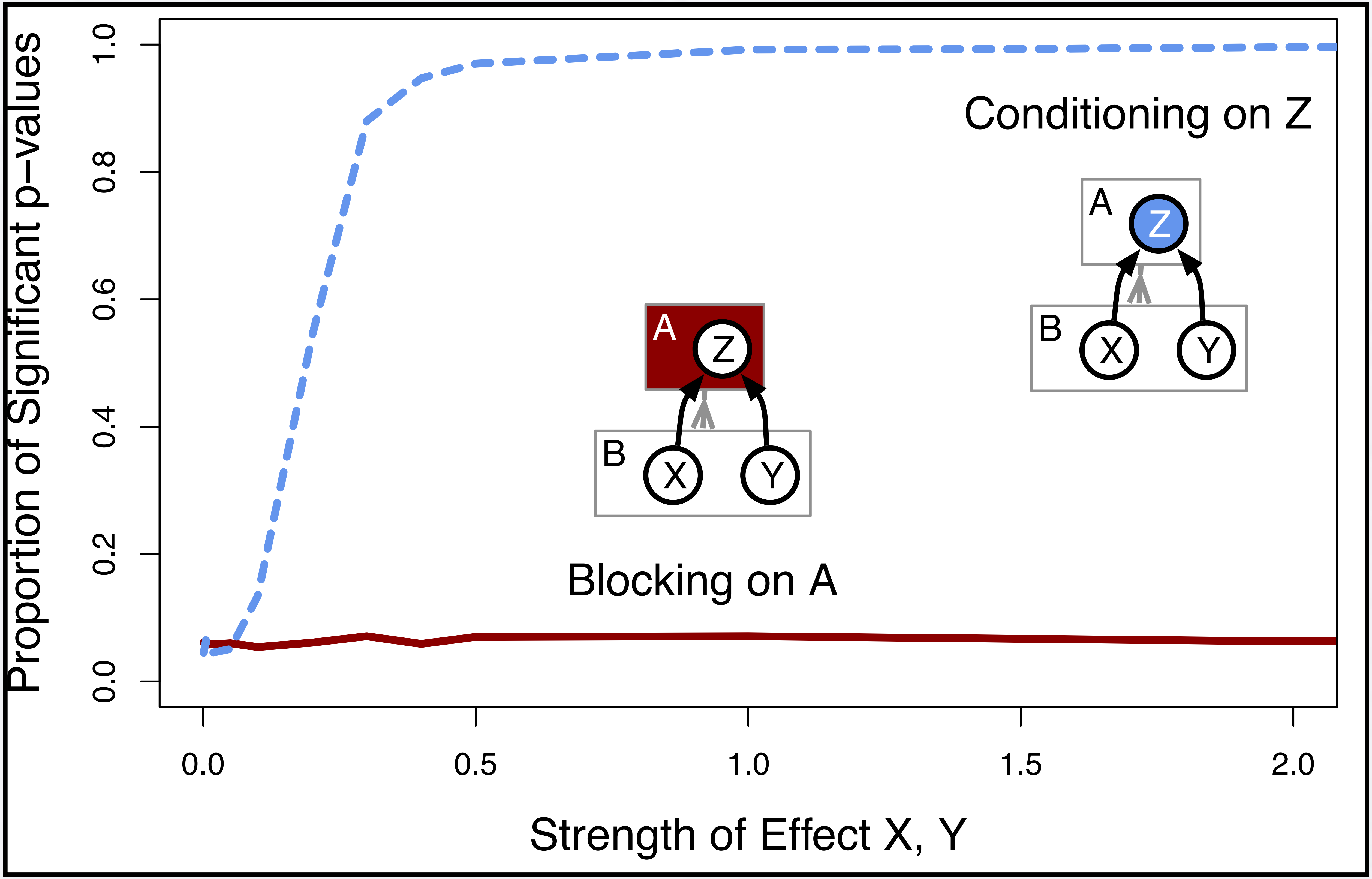
p=0.009

p=0.033

Latent common causes



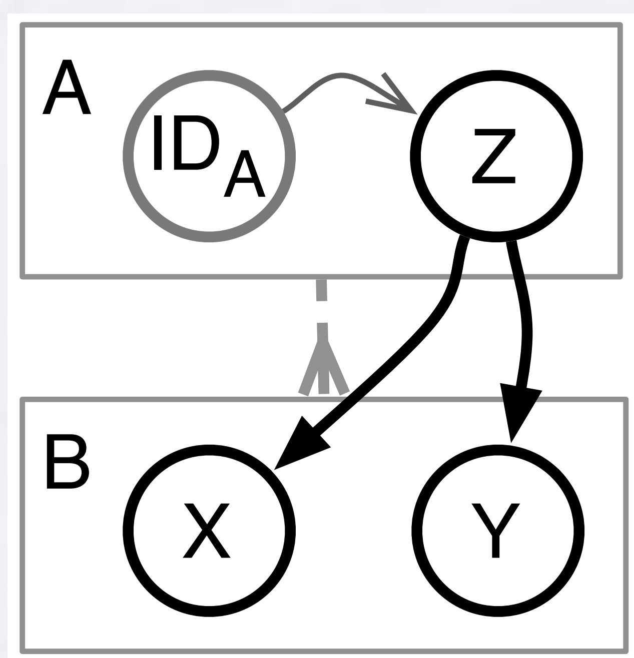
Common effects

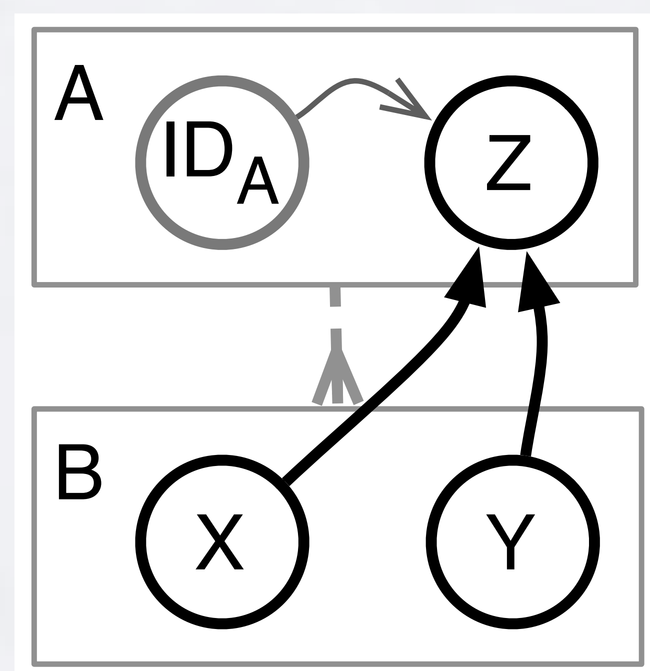


D-separation

Let $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{W}$ be three disjoint sets of vertices in DAG G.
Let $\mathrm{Det}(\mathbf{V})$ be the set of all variables determined by $\mathbf{V}$.
Then, $\mathbf{X}$ and $\mathbf{Y}$ are ***d*-separated** by $\mathbf{W}$ if and only if for all undirected paths P between $\mathbf{X}$ and $\mathbf{Y}$ either

(1) $\exists v \in \text{colliders}(P)$ such that $v \wedge \text{descendants}(v) \notin \mathbf{W}$ or
(2) $\exists v \in \text{noncolliders}(P)$ such that $v \in \mathrm{Det}(\mathbf{W})$.





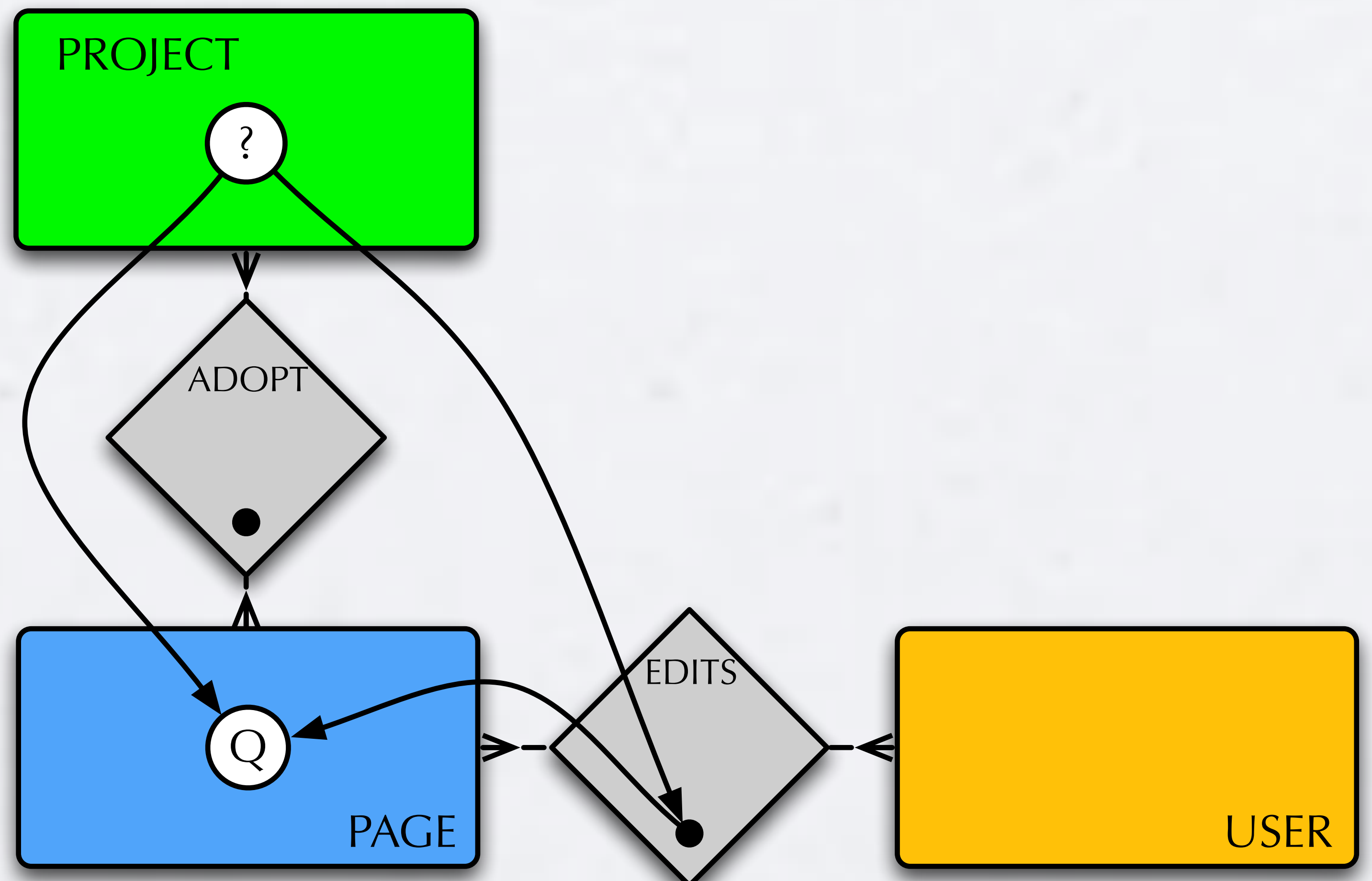
Example of relational blocking

Domain: Wikipedia

Question: Do *"many eyes"* cause *quality*?

Treatment: Number of distinct editors

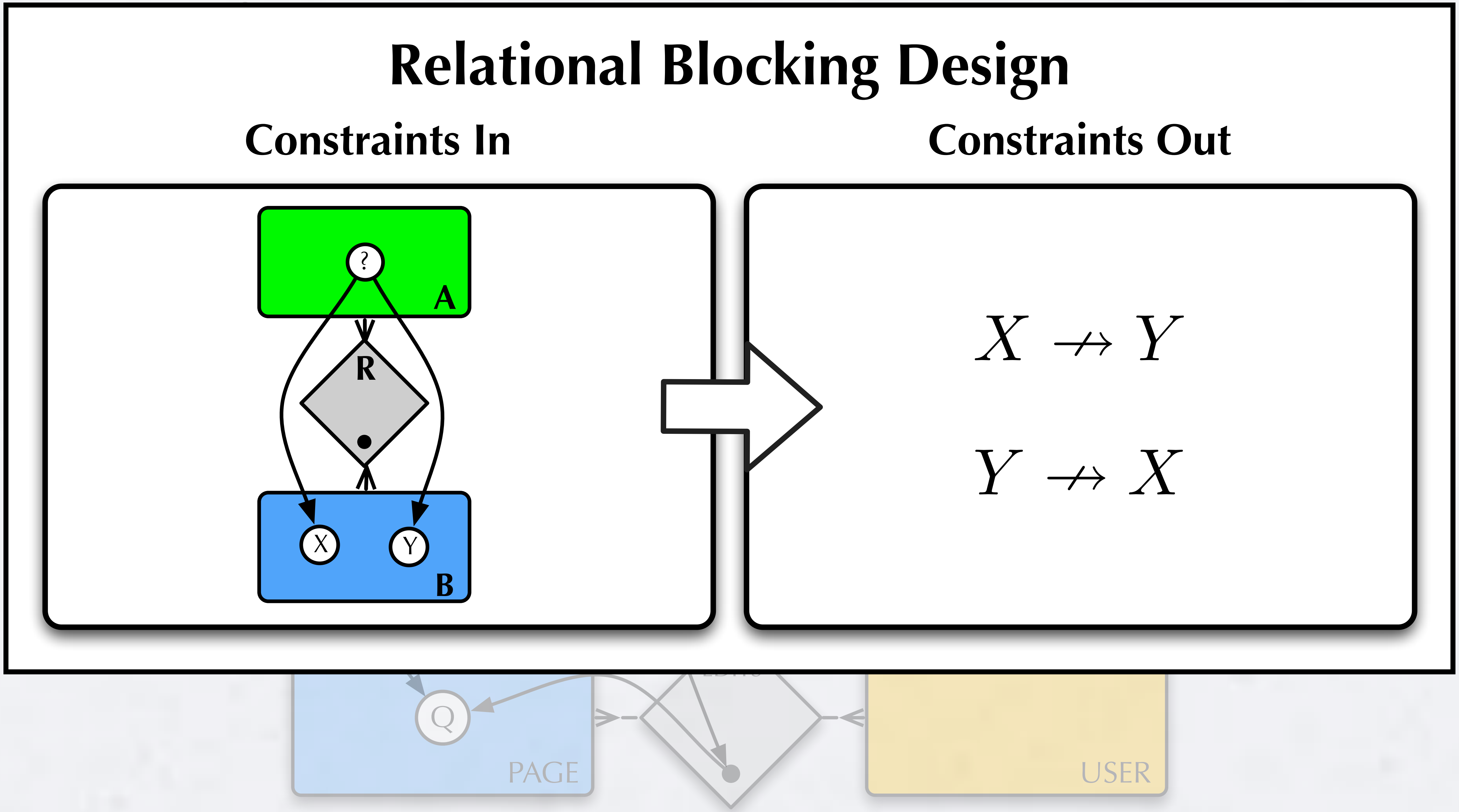
Outcome: Assessed quality



(Rattigan, Maier, Jensen 2010)

Example of relational blocking

Domain: Wikipedia



(Rattigan, Maier, Jensen 2010)

QEDs

Quasi-Experimental Design

- Techniques developed and used by social scientists
  - Provide *causal conclusions*
  - Devise *local hypothesis tests*
  - Enabled by *temporal and relational representation*
  - Lack a *formalization*

- **Relational blocking**: twin study, matching design
- **Temporal blocking**: interrupted time series design, non-equivalent control group design

(Cook & Campbell 1979;
Shadish, Cook, Campbell 2002)

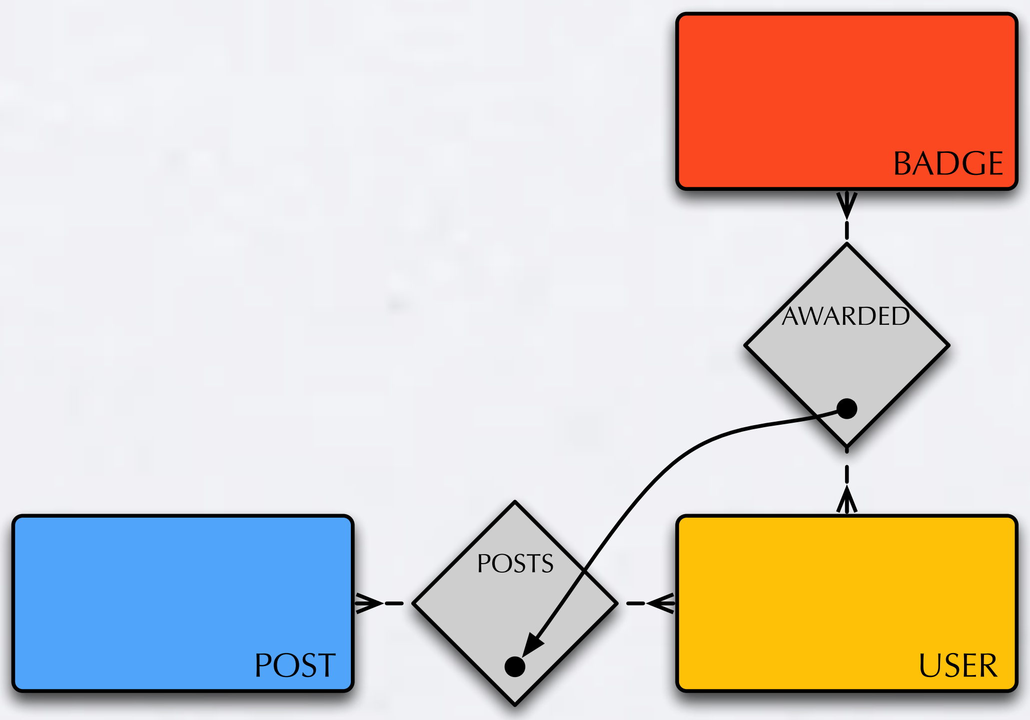
Example of temporal blocking

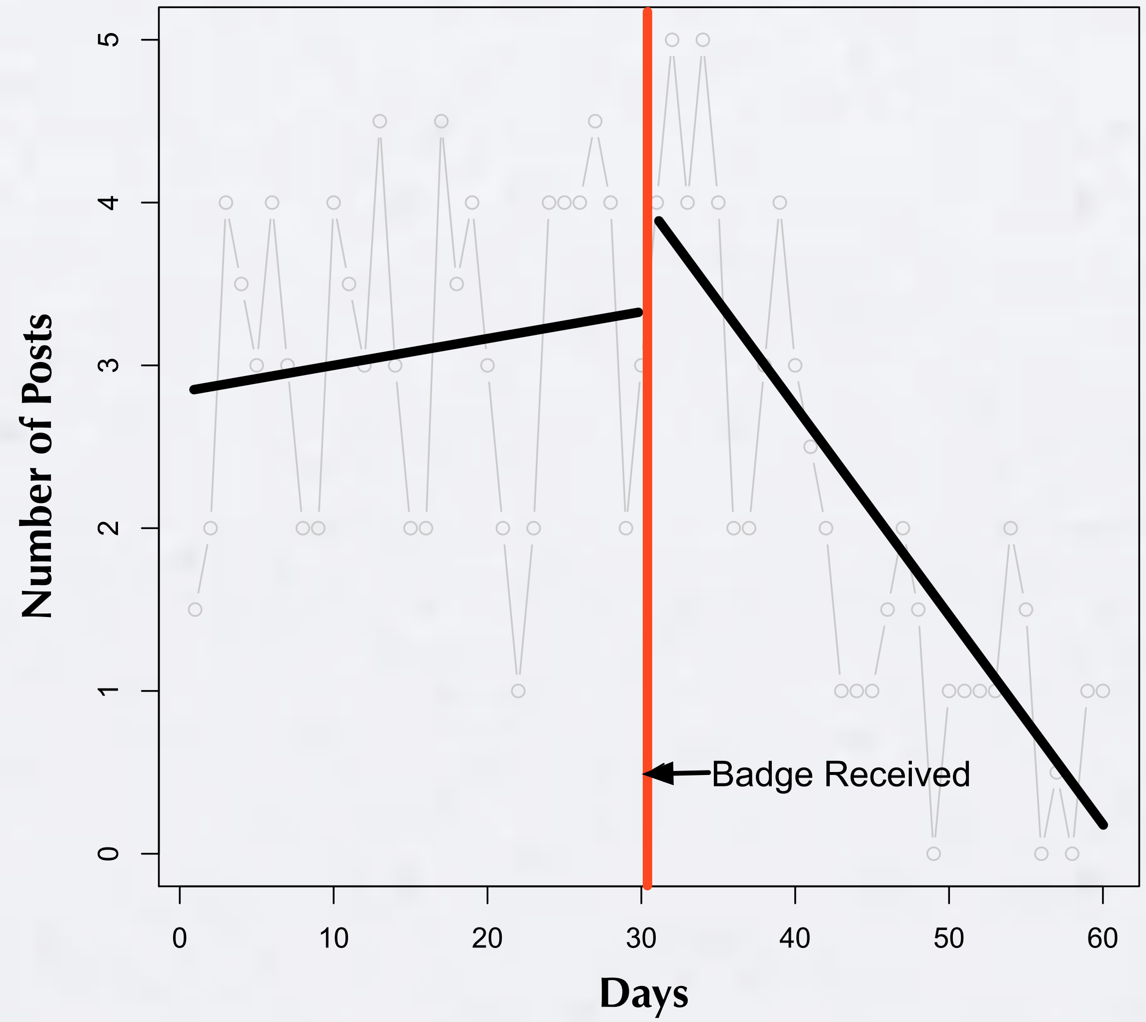
Domain: Stack Overflow

Question: Do *badges* influence *participation*?

Treatment: User receives "epic" badge

Outcome: Posting frequency over time





(Oktay, Taylor, Jensen 2010)

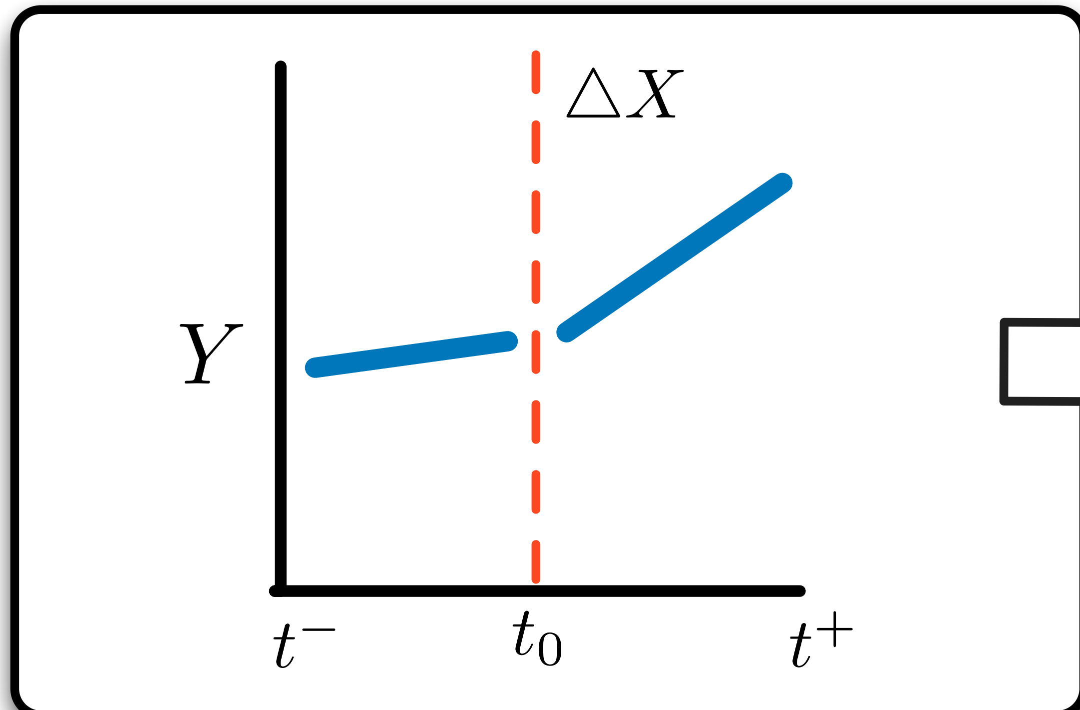
Example of temporal blocking

Domain: Stack Overflow

Question: Do *badges* influence participation?

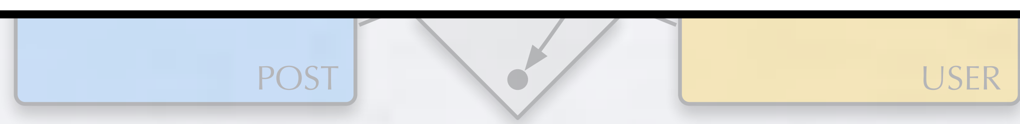
Temporal Blocking Design

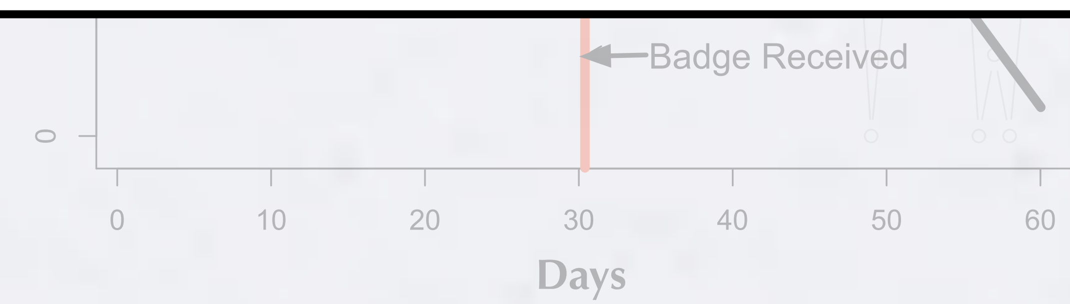
Constraints In



Constraints Out

$$X \to Y$$





(Oktay, Taylor, Jensen 2010)

Thank you!

Questions?

maier@cs.umass.edu