Causal Relation Learning

Babak Salimi, Harsh Parish, Moe Kayali, Sudeepa Roy, Lise Getoor, and Dn Suciu

Presented by: Vishal C. & Pratyoy D.

CS 276 , W'24

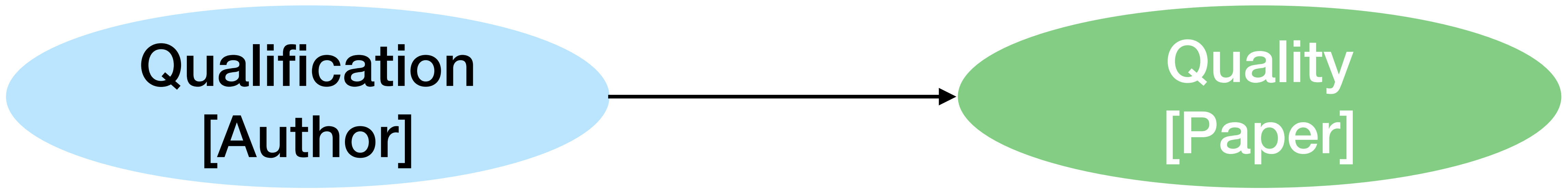
13 March, 2024

Goal

Causal Inference in Relational Databases

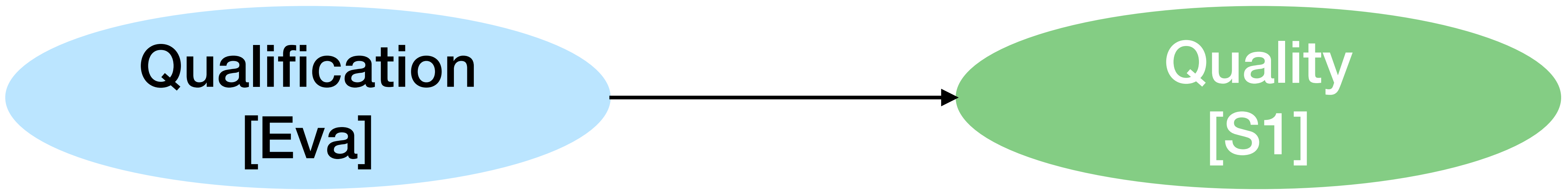
Why is Regular Causal Models not sufficient?

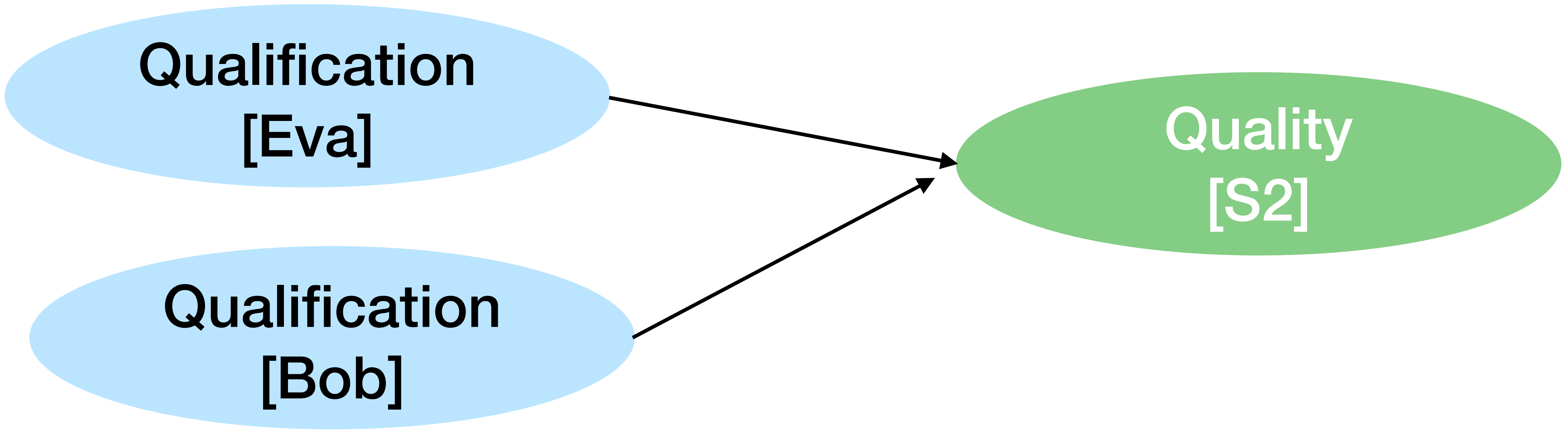
Data is mostly not homogenous



Pearl’s Causal Model

But papers may have different number of authors who impact the quality differently





Driving Use Case

Running Example

- A relational database of conference paper submissions
- Ask “Does single blind conferences favour authors from prestigious institutes?”
  - SQL can show correlation, but not causation —need Causal Learning

Authors		
person	prestige	qualification (h-index)
Bob	1	50
Carlos	0	20
Eva	1	2

Submissions	
sub	score
s1	0.75
s2	0.4
s3	0.1

Authorship	
person	sub
Bob	s1
Eva	s1
Eva	s2
Eva	s3
Carlos	s3

Submitted	
sub	conf
s1	ConfDB
s2	ConfAI
s3	ConfAI

Conferences	
conf	blind
ConfDB	Single
ConfAI	Double

Introducing CaRL

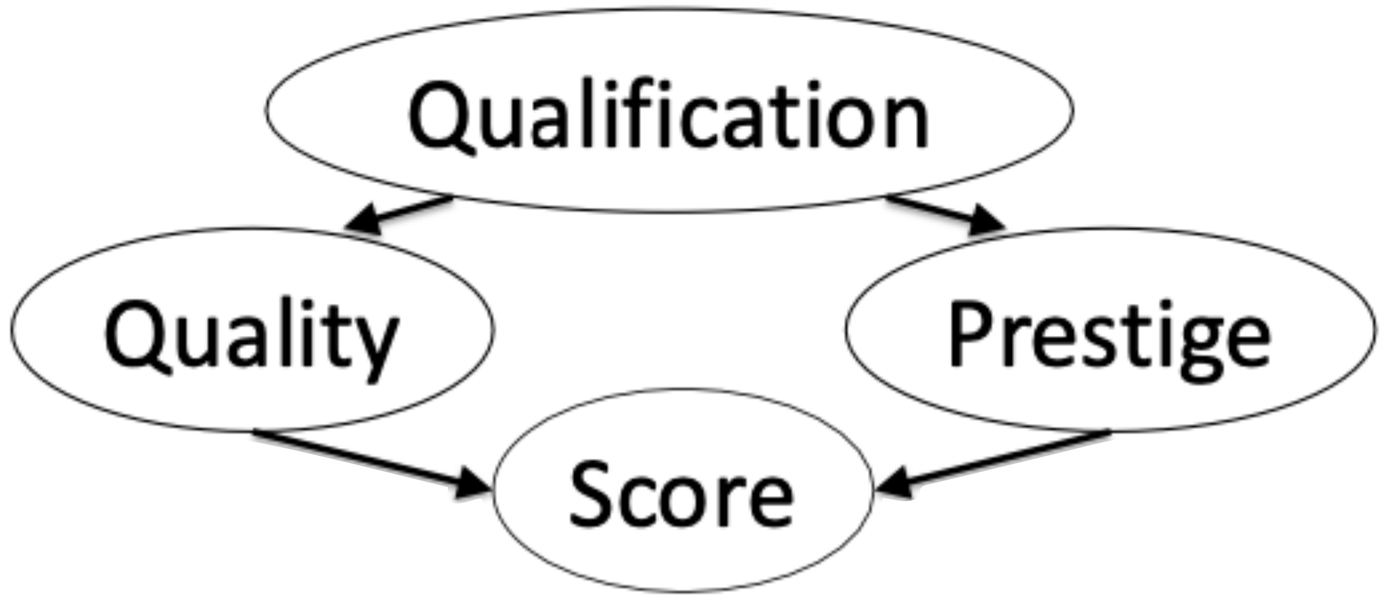
Main Contributions

- A declarative language **CaRL (Causal Relational Language)**
  - representing causal background knowledge and assumptions in relational domains
- Define semantics for complex causal relational-queries
  - treatment units and outcome units might be heterogeneous
- An algorithm for answering causal queries from the given relational data
  - Performing a static analysis of the causal query

Components of CaRL

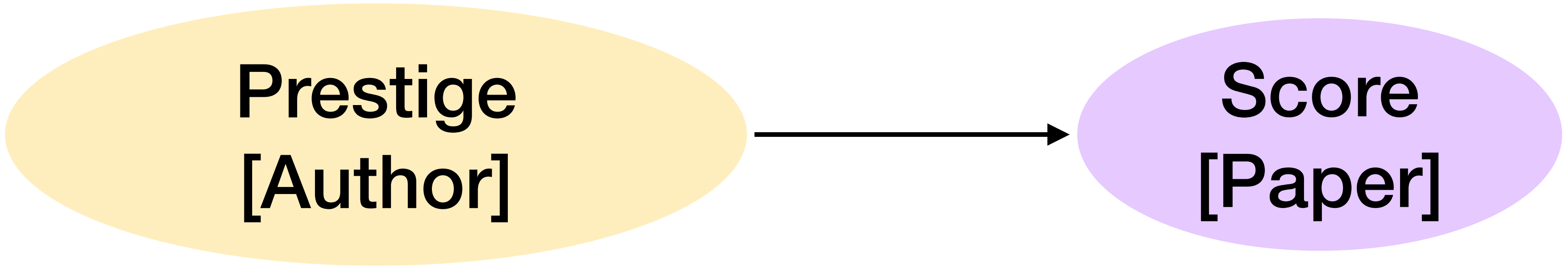
Overview

Relational DB

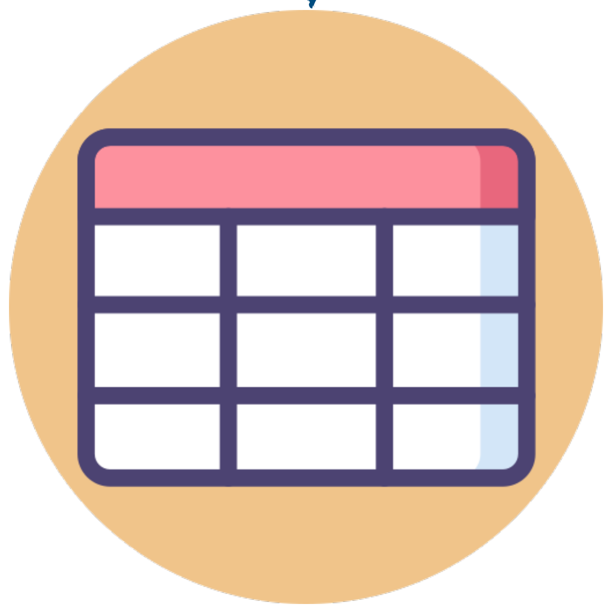


Causal Graph

(From Intuition)



Causal Query



Flat Table Structure

Causal Effect Estimates

Relational Model

Extending the Entity-Relation Model

- Schema $S = (P, A)$
  - P = Entities(E) ∪ Relationships(R)
  - A is the set of Attribute Functions (or Attributes)
- Examples of **Entities**
  - Author (Bob), Author (Eva), Submission (P1), Submission(P2)
- Examples of **Relationships**
  - Authorship(Bob, P1) , Authorship(Eva, P1), Authorship (Eva,P2)

Author

Author	Prestige	Qualification
Bob	1	25
Eva	0	2

Submission

PaperId	Score	Quality
P1	0.75	1
P2	0.25	0

Authorship

PaperID	AuthoredBy
P1	Bob
P2	Eva
P1	Eva

Attribute Functions

- $A[X]$ where A is an observable attribute
- Examples of Attribute Functions:-
  - Qualification[Bob], Prestige[Bob]
- Some attributes are observable while others aren't. ($\boldsymbol{A}_{obs} \subset \boldsymbol{A}$)
- Attributes can be mutable but Entities and Relationships are not!

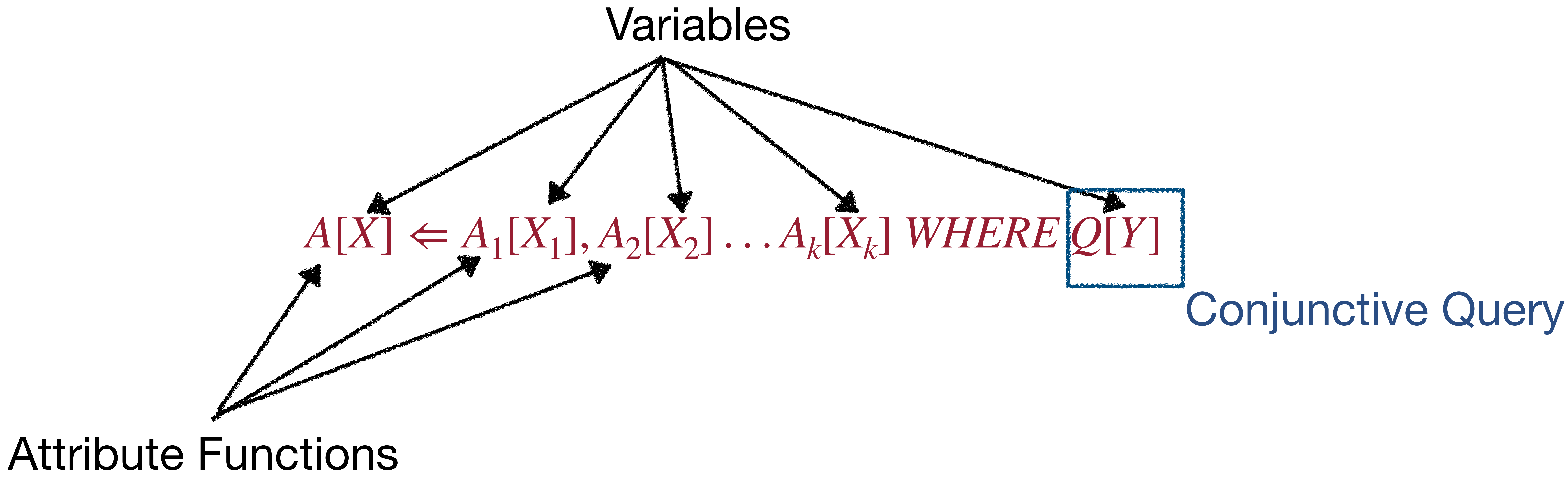
Author

Person	**Prestige**	**Qualification**
Bob	1	25
Eva	0	2

Relational Causal Rules

Normal Form

- Background Knowledge can be modeled using relational causal rules.



Examples of Causal Rules

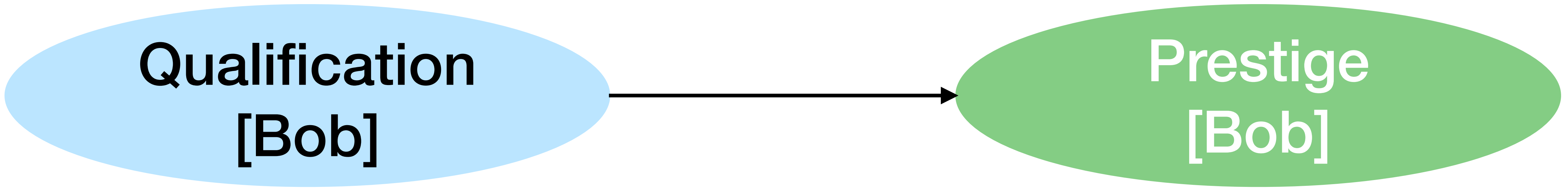
- $PRESTIGE[A] \Leftarrow Qualification[A]$ WHERE $Person[A]$
  - Qualification of a person causally affects his or her institutions' prestige
- $Quality[S] \Leftarrow Prestige[A], Qualification[A]$ WHERE $Author[A, S]$
  - Quality of a submission is affected by its authors' qualifications and prestige

Instantiated Rules

- Causal Rules which have been instantiated with database constants
  - $PRESTIGE[A] \Leftarrow Qualification[A]$ WHERE $Person[A]$
    - $PRESTIGE[Bob] \Leftarrow Qualification[Bob]$

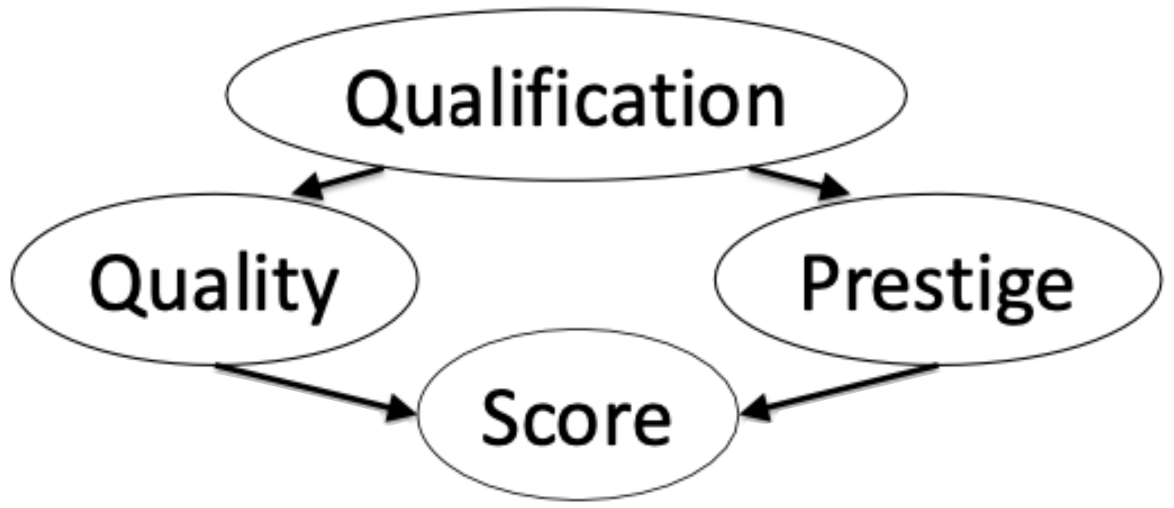
Person	Prestige	Qualification
Bob	1	50
Eva	0	75

- A causal graph G can be constructed from the set of Instantiated Rules
- For every instantiated rule, we have an edge



Relational Causal Graph

Extension of Pearl's Causal Graph



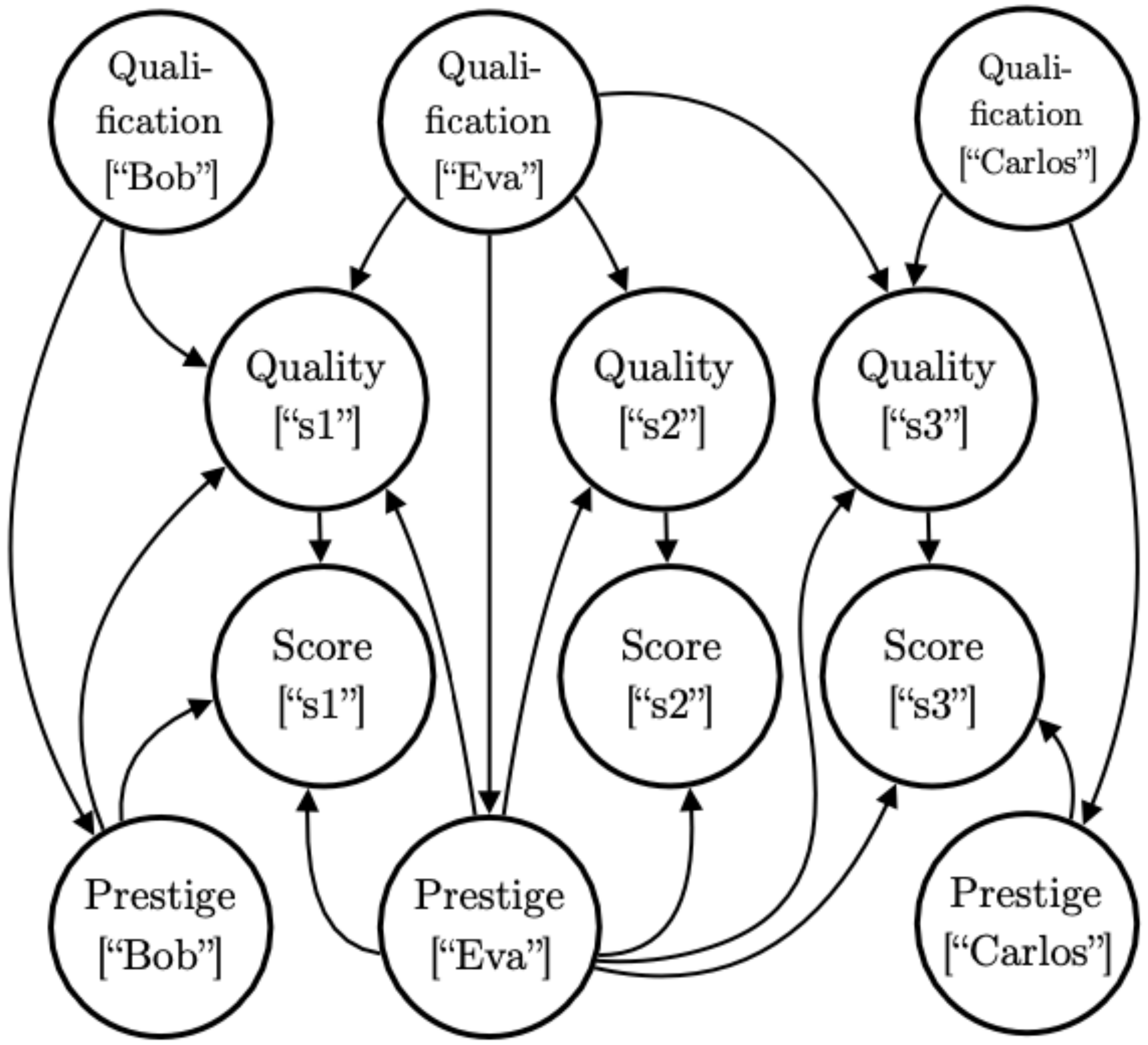
Authors		
person	prestige	qualification (h-index)
Bob	1	50
Carlos	0	20
Eva	1	2

Submissions	
sub	score
s1	0.75
s2	0.4
s3	0.1

Authorship	
person	sub
Bob	s1
Eva	s1
Eva	s2
Eva	s3
Carlos	s3

Submitted	
sub	conf
s1	ConfDB
s2	ConfAI
s3	ConfAI

Conferences	
conf	blind
ConfDB	Single
ConfAI	Double



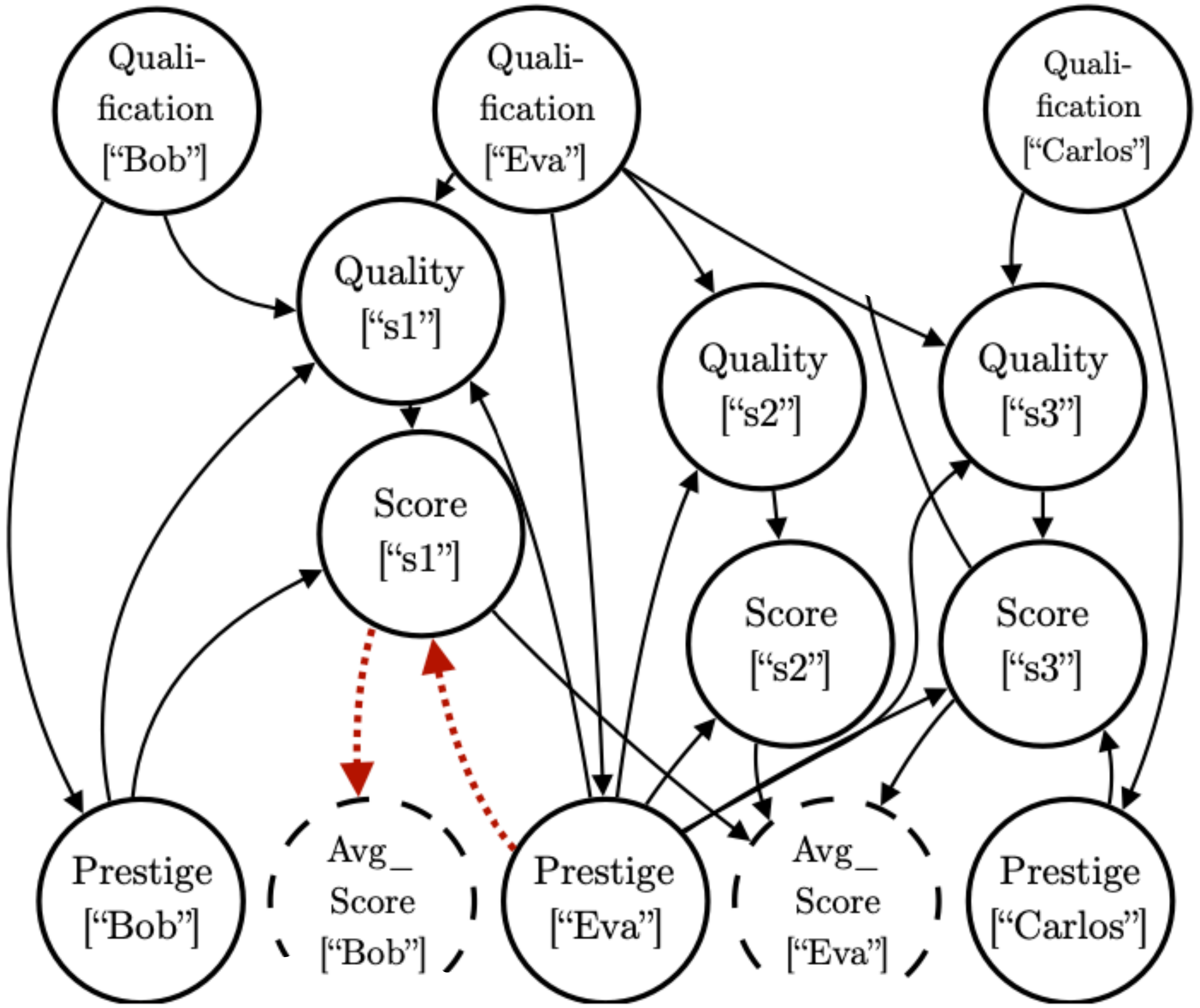
- Multiple nodes for every "type" of unit
  - Score: $Score[s1], Score[s2]$ - one for each submission
- Relation Causal graph defines a joint probability
  - $Pr([A_x] \mid Pa[A_x])$
  - with one conditional probability on each ground rule

Aggregated Rules

- Extend set of attribute functions A with new aggregated functions using aggregated rules
- $AGG\_A[W] \Leftarrow A[X]$ WHERE $Q[Z]$
- The new aggregated attribute functions AGG_A are included in the extended attribute functions A
- Similar to relational causal rules, aggregated rules define a set of grounded rules with corresponding vertices and edges in the relational causal graph
  - However, instead of a conditional probability distribution, a deterministic function $AGG(Pa(AGG\_Y[w]))$ will be associated with each $AGG\_Y[w]$

Example of Aggregated Rules

- $AVG\_SCORE[A] \Leftarrow SCORE[S]$ WHERE $AUTHOR[A, S]$
- We can construct an Extended relational causal graph with aggregated attribute AVG_Score[A]
    - The directed path from relational peer Eva's prestige to average score of Bob is highlighted



Causal Query Language in CaRL

Supported Queries

- Compare papers' scores in two hypothetical worlds in which all authors are and are not affiliated with prestigious institutions
    - $Score[S] \Leftarrow Prestige[A]?$
- Compute the treatment effect of the prestige of authors on the average score received by author
    - $AGG\_Y[X'] \Leftarrow T[X]?$
- Computes values for (i) isolated (an author's prestige), (ii) relational (his/her coauthor's prestige), and (iii) overall (all authors' prestige) effect of prestige on a submission's score.
    - $Y[X'] \Leftarrow T[X]?\ WHEN\ \langle cnd \rangle\ PEERS\ TREATED$

Semantics For Relational Causal Analysis

Complexities in a Relational Causal Graph

- Probability distribution given by $Pr(X \mid \boldsymbol{Pa}(X))$
- Standard Causal Graphs
  - Unknown but can be estimated from available data
  - Fixed number of nodes and edges
- Relational Causal Graph
  - Unknown but can be estimated from available data
  - **Number of nodes depends on instantiations**

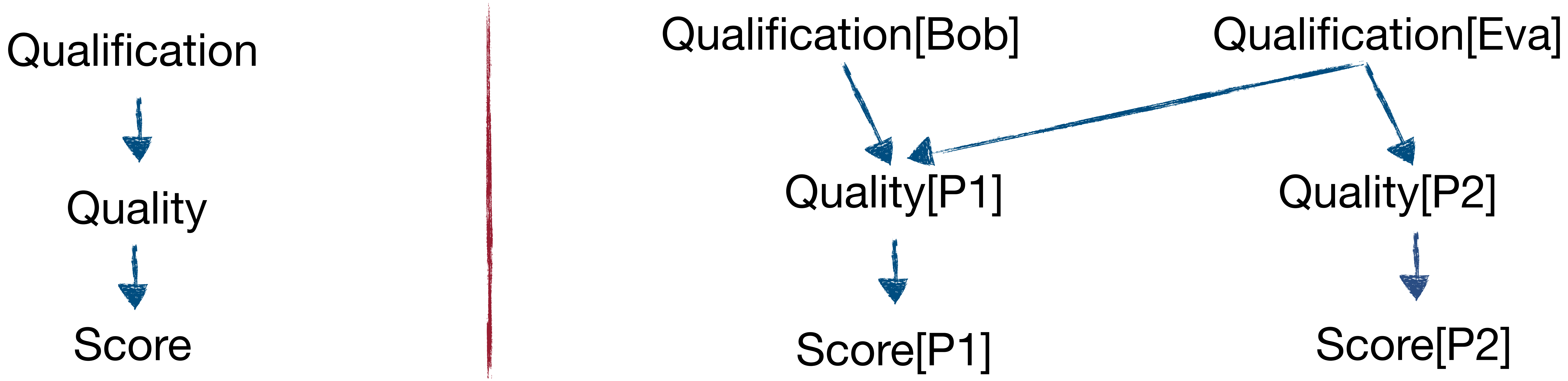
Structural Homogeneity Assumption

Example: Number of nodes depend on instantiations

Author	Prestige	Qualification
Bob	1	50
Eva	0	75

PaperId	Score	Quality
P1	0.75	1
P2	0.25	0

PaperID	AuthoredBy
P1	Bob
P2	Eva
P1	Eva



Embedding Functions

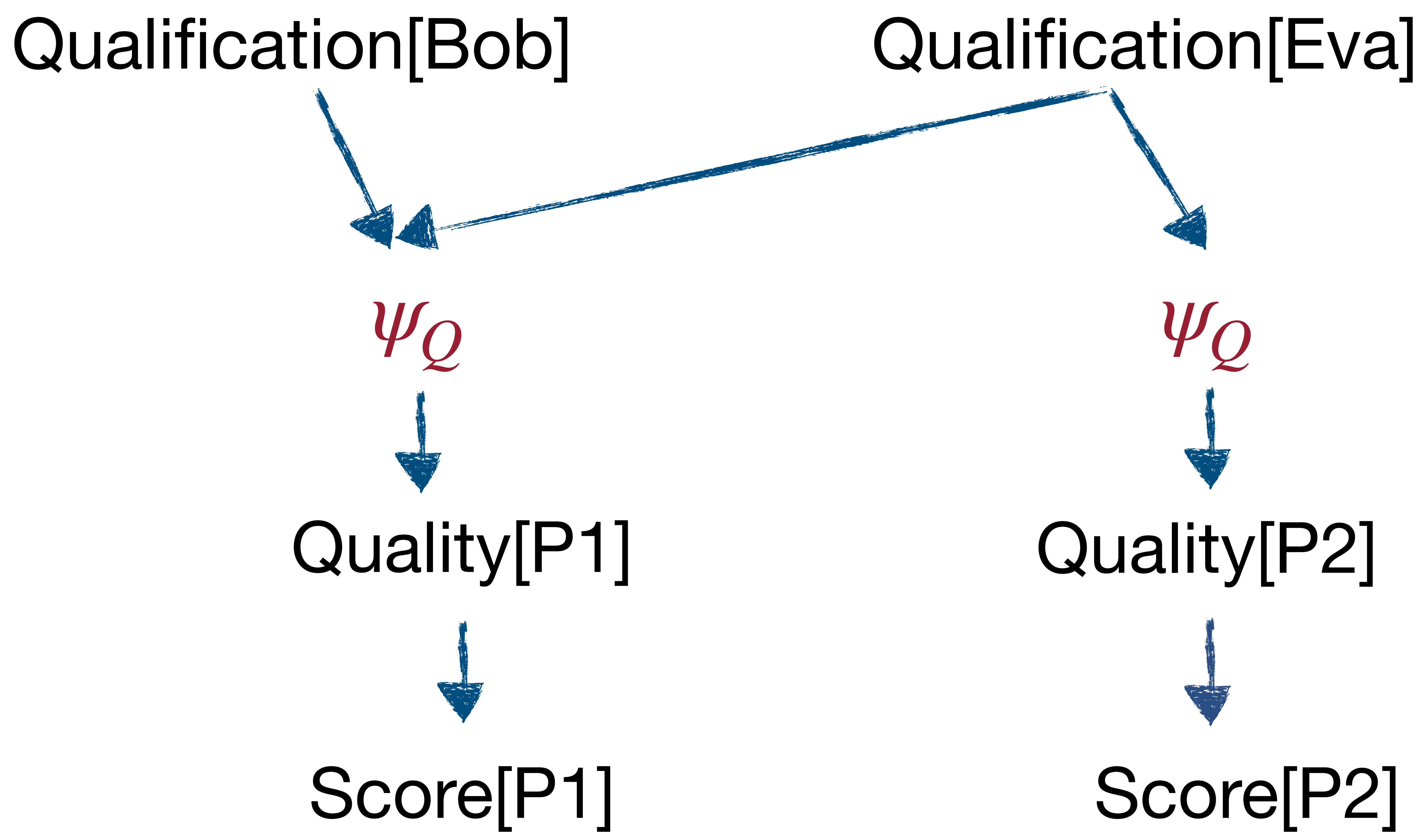
Structural Homogeneity Assumptions

<u>Low dimensional Vector</u>

Mean

Median

Padding



Structural Homogeneity Assumption

Redefining Probability Distributions

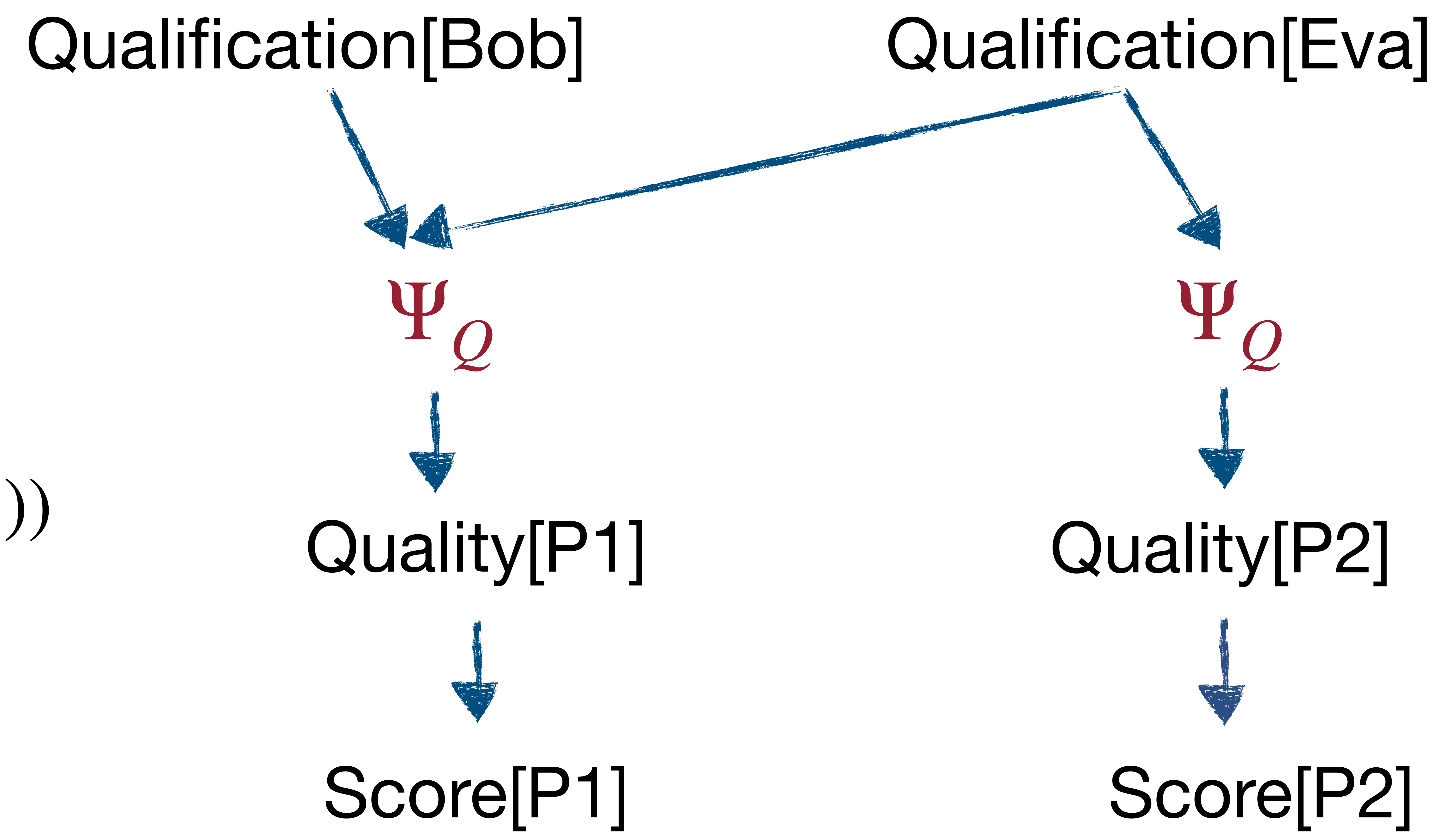
$$\Pr(A[\boldsymbol{x}] \mid \Psi^A(\boldsymbol{Pa}(A[\boldsymbol{x}])))$$

Ψ^A **Collection of mappings that projects parents of $A[x]$ into a low-dimension vector with fixed dimensionality for all $A[x]$**

Structural Homogeneity Assumption

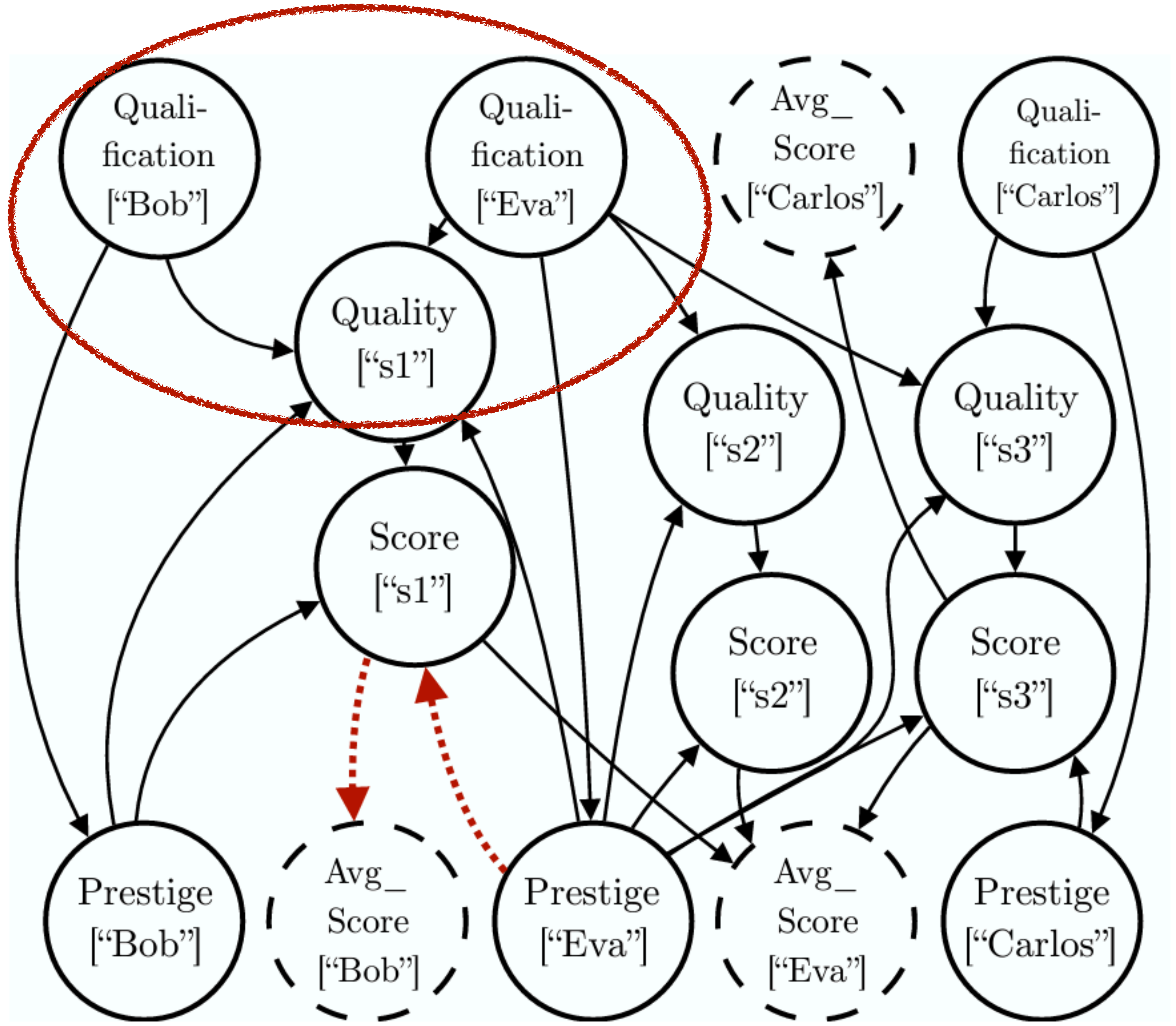
Redefining Probability Distributions

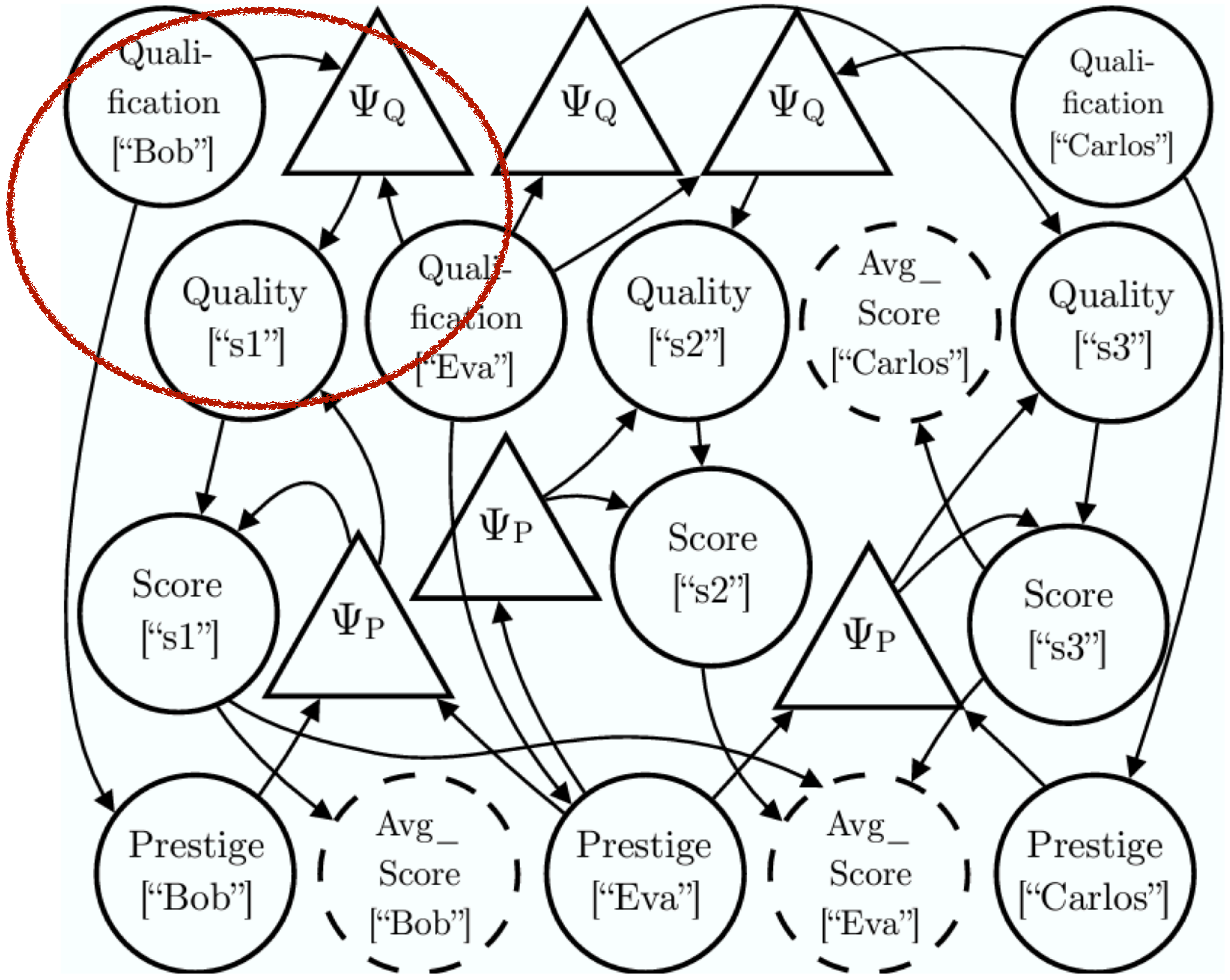
$$\Pr(\boldsymbol{A}) = \Pi_{A[\boldsymbol{x}]\in\boldsymbol{A}} \Pr(A[\boldsymbol{x}] \mid \Psi^{A}(\mathrm{Pa}(A[\boldsymbol{x}])))$$



Embeddings

Example





Treated And Response Units

Covariate Detection

Real Binary Values

- Treatment Attribute Function $T[\boldsymbol{X}]$
- Response Attribute Function $Y[\boldsymbol{X}']$

Example: Want to find effects of author's Prestige on submission scores

$Prestige[A]$

$Score[P]$

Treated And Response Units

Covariate Detection

- Set of treated units: $\mathbb{U}_T = \{x_1, x_2, \ldots .\}$
- Binary vector: $\vec{t} = (t_1, t_2, \ldots)$
- Intervention do$(T(x_i) = t_i)$ on all related units x_i

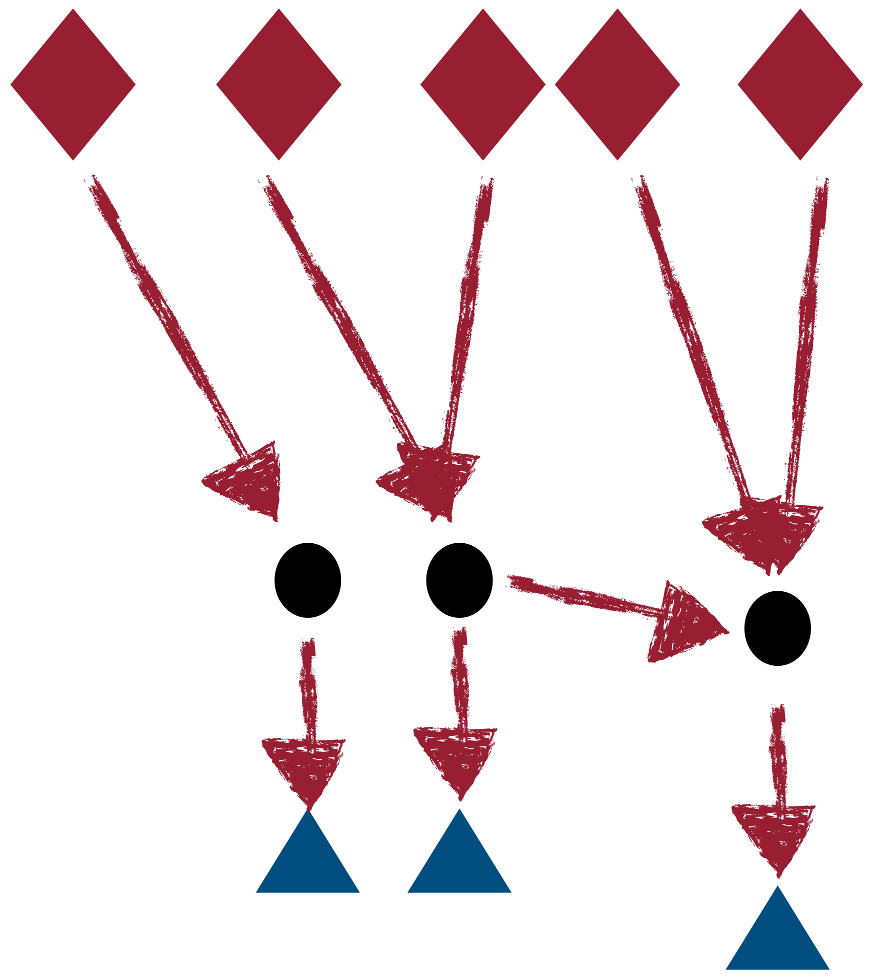
Example: Set all author's prestige to 1 (they are form prestigious schools)

$$\vec{1} = (1,1,\ldots) \longrightarrow Prestige[A] \longrightarrow Score[P]$$

Relational Peers

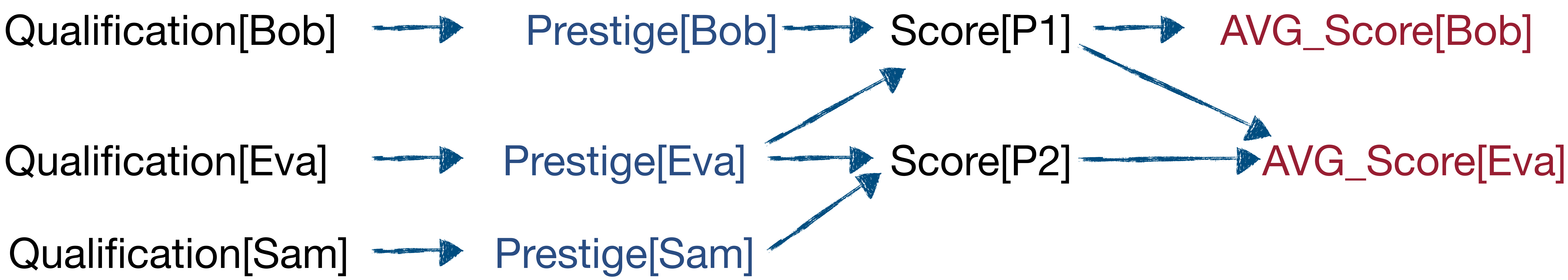
Treatment and Response

- Given treated attribute function $T[\boldsymbol{X}]$ and response attribute function $Y[\boldsymbol{X}]$
- Relational Peers of $x \in \mathbb{U}_{(T,Y)}$ as a set of units $\mathbb{P}(x) = \mathbb{U}_{(T,Y)} - \{x\}$
  - s.t. for each $p \in \mathbb{P}(x)$ there is a path from $T[\boldsymbol{p}]$ to $Y[x]$ in G



Expected Response Unit On Being Treated

Covariate Detection



Treatment: $Prestige[X]$

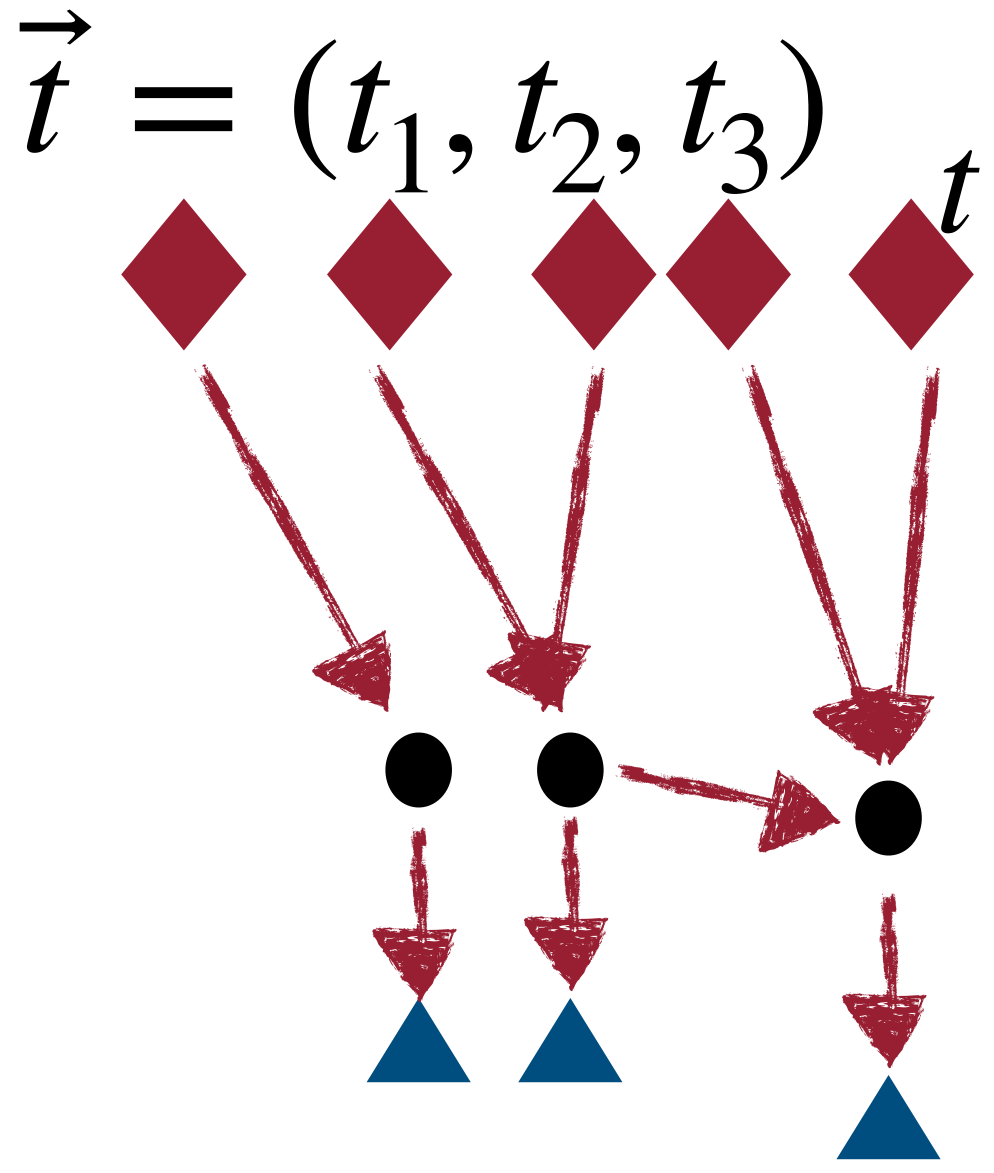
Response: $AVG\_Score[X]$

$\mathbb{P}(Bob) = \{Eva\}$ — $Prestige[Eva]$ path to $AVG\_Score[Bob]$

$\mathbb{P}(Eva) = \{Bob, Sam\}$ — $Prestige[Bob]$ path to $AVG\_Score[Eva]$; $Prestige[Sam]$ path to $AVG\_Score[Eva]$

Expected Response Unit On Being Treated

Covariate Detection



$T[\boldsymbol{x}]$ recieves treatment t

$\mathbb{P}(\boldsymbol{x})$ recieves treatment $\vec{t}$

$$Y_{\boldsymbol{x}}(t, \vec{t}) = \mathbb{E}[Y[\boldsymbol{x}] \mid \mathrm{do}(T(x) = t), \mathrm{do}(T[\mathbb{P}(x)] = \vec{t})]$$

Q1: Average Treatment Effect

Semantics

$$Y[\boldsymbol{X}'] \Leftarrow T[\boldsymbol{X}]?$$

Score[P1]⇐Prestige[Bob]?

$$ATE(T, Y) = \sum_{\boldsymbol{x}' \in \mathbb{U}_Y} \frac{1}{m} \; \mathbb{E}[Y[\boldsymbol{x}'] \mid \text{do}(T[\mathbb{U}_T]) = \vec{0}) - \mathbb{E}[Y[\boldsymbol{x}'] \mid \text{do}(T[\mathbb{U}_T]) = \vec{1})$$

Compare under two scenarios: with intervention and without

Q2: Aggregated Responses Queries

Semantics

$$AGG\_Y[\boldsymbol{X}'] \Leftarrow T[\boldsymbol{X}]?$$

AGG_Score[S]⇐Prestige[A]?

$$ATE(T, AGG\_Y) = \sum_{\boldsymbol{x}' \in \mathbb{U}_Y} \frac{1}{m} \mathbb{E}[AGG\_Y[\boldsymbol{x}'] \mid \mathrm{do}(T[\mathbb{U}_T]) = \vec{0}) - \mathbb{E}[AGG\_Y[\boldsymbol{x}'] \mid \mathrm{do}(T[\mathbb{U}_T]) = \vec{1})$$

Q3: Average Isolated Effect (AIE)

Semantics

$\boldsymbol{x}$ recieves treatment t

$\mathbb{P}(\boldsymbol{x})$ recieves treatment $\vec{t}$

$\boldsymbol{x}$ recieves treatment t'

$\mathbb{P}(\boldsymbol{x})$ recieves treatment $\vec{t}$

$$ATE(t; t' \mid \vec{t}) = \frac{1}{n} \sum_{\boldsymbol{x}' \in \mathbb{U}_{(T,Y)}} Y_X(t, \vec{t}) - Y_X(t', \vec{t})$$

Q4: Average Relational Effect (ARE)

Semantics

$\boldsymbol{x}$ recieves treatment t

$\mathbb{P}(\boldsymbol{x})$ recieves treatment $\vec{t}$

$\boldsymbol{x}$ recieves treatment t

$\mathbb{P}(\boldsymbol{x})$ recieves treatment $\vec{t}'$

$$ARE(\vec{t}, \vec{t}' \mid t) = \frac{1}{n} \sum_{\boldsymbol{x}' \in \mathbb{U}_{(T,Y)}} Y_X(t, \vec{t}) - Y_X(t, \vec{t}')$$

Q5: Average Overall Effect (ARE)

Semantics

$\boldsymbol{x}$ recieves treatment t

$\mathbb{P}(\boldsymbol{x})$ recieves treatment $\vec{t}$

$\boldsymbol{x}$ recieves treatment t'

$\mathbb{P}(\boldsymbol{x})$ recieves treatment $\vec{t}'$

$$ATE(t, \vec{t}; t', \vec{t}') = \frac{1}{n} \sum_{\boldsymbol{x}' \in \mathbb{U}_{(T,Y)}} Y_X(t, \vec{t}) - Y_X(t', \vec{t}')$$

Relationships between Average Effects

Semantics

$\boldsymbol{x}$ recieves treatment t

$\mathbb{P}(\boldsymbol{x})$ recieves treatment $\vec{t}$

$\boldsymbol{x}$ recieves treatment t'

$\mathbb{P}(\boldsymbol{x})$ recieves treatment $\vec{t}'$

$$ATE(t, \vec{t}; t', \vec{t}') = AIE(t, t' \,|\, \vec{t}) + ARE(\vec{t}, \vec{t}' \,|\, t')$$

$$= AIE(t, t' \,|\, \vec{t}') + ARE(\vec{t}, \vec{t}' \,|\, t)$$

Answering Causal Queries

CaRL

- Covariate detection
  - identify a sufficient set of covariates that should be adjusted for to remove confounding effects
- Covariate adjustment
  - the data is transformed into a flat, single table format so that causal inference can be performed using standard methods.

Covariate Detection
CaRL

- Recall $Y_{\boldsymbol{x}}(t, \vec{t}) = \mathbb{E}[Y[\boldsymbol{x}] \mid \text{do}(T(x) = t), \text{do}(T[\mathbb{P}(x)] = \vec{t})]$
- Estimate quantities of the form $\mathbb{E}[Y[\boldsymbol{x}] \mid \text{do}(T(x) = t)$
  - Graphical criterion to select a sufficient set of covariates from a G

Relational Adjustment Formula

Intuition

$$\mathbb{E}[Y[\boldsymbol{x}] \mid \text{do}(T(\mathbb{S}) = \overrightarrow{t_{\mathbb{S}}})] = \sum_{z \in Dom(\boldsymbol{Z})} \mathbb{E}[Y[\boldsymbol{x}'] \mid \boldsymbol{Z} = \boldsymbol{z}, T([\mathbb{S}'] = \overrightarrow{t_{\mathbb{S}'}}]\text{Pr}(\boldsymbol{Z} = \boldsymbol{z})$$

always sufficient to condition for the ‘parents’ of treated units as they separate effects from the rest of the graph ensuring independence.

$$[Y[\boldsymbol{x}'] \perp\!\!\!\perp (\bigcup_{\boldsymbol{x} \in \mathbb{S}} \boldsymbol{Pa}(T[\boldsymbol{x}]) \,|_G (\bigcup_{x \in \mathbb{S}} T[\boldsymbol{x}], \boldsymbol{Z})$$

Relational Adjustment Formula

Theorem

Given: relational graph G, treatment T, response Y, set $\mathbb{S}$ of treatment units with the treatment assignment $\overrightarrow{t_{\mathbb{S}}}$

$$\mathbb{E}[Y[\boldsymbol{x}] \mid \mathrm{do}(T(\mathbb{S}) = \overrightarrow{t_{\mathbb{S}}})] = \sum_{\mathrm{Z} \in Dom(\boldsymbol{Z})} \mathbb{E}[Y[\boldsymbol{x}'] \mid \boldsymbol{Z} = \boldsymbol{z}, T([\mathbb{S}'] = \overrightarrow{t_{\mathbb{S}'}}]\Pr(\boldsymbol{Z} = \boldsymbol{z})$$

where $\mathbb{S}' \subseteq \mathbb{S}$ is such that, for each unit $x \in \mathbb{S}'$, there exists a directed path from $T[\boldsymbol{x}]$ to the node $Y[\boldsymbol{x}']$ in G, and $\boldsymbol{Z}$ is set of nodes in G corresponding to the groundings of observed attribute functions A_{Obs} such that

$$[Y[\boldsymbol{x}'] \perp\!\!\!\perp (\bigcup_{\boldsymbol{x} \in \mathbb{S}} \boldsymbol{Pa}(T[\boldsymbol{x}]) \,|_G (\bigcup_{x \in \mathbb{S}} T[\boldsymbol{x}], \boldsymbol{Z})$$

Covariate Adjustment

Overview

- When the set of confounding covariates $\boldsymbol{Z}$ has high dimensionality
  - Estimating the conditional expectation is hard. One for each peer!

$$\sum_{\mathbf{Z} \in Dom(\boldsymbol{Z})} \mathbb{E}[Y[\boldsymbol{x}'] \mid \boldsymbol{Z} = \boldsymbol{z}, T([\mathbb{S}'] = \overrightarrow{t_{\mathbb{S}'}}]\text{Pr}(\boldsymbol{Z} = \boldsymbol{z})$$

- The causal queries need to compute averages across all response units

Evaluation

What to evaluate?

1. End to end performance
2. Correctness
3. How does embedding affect results?

Datasets

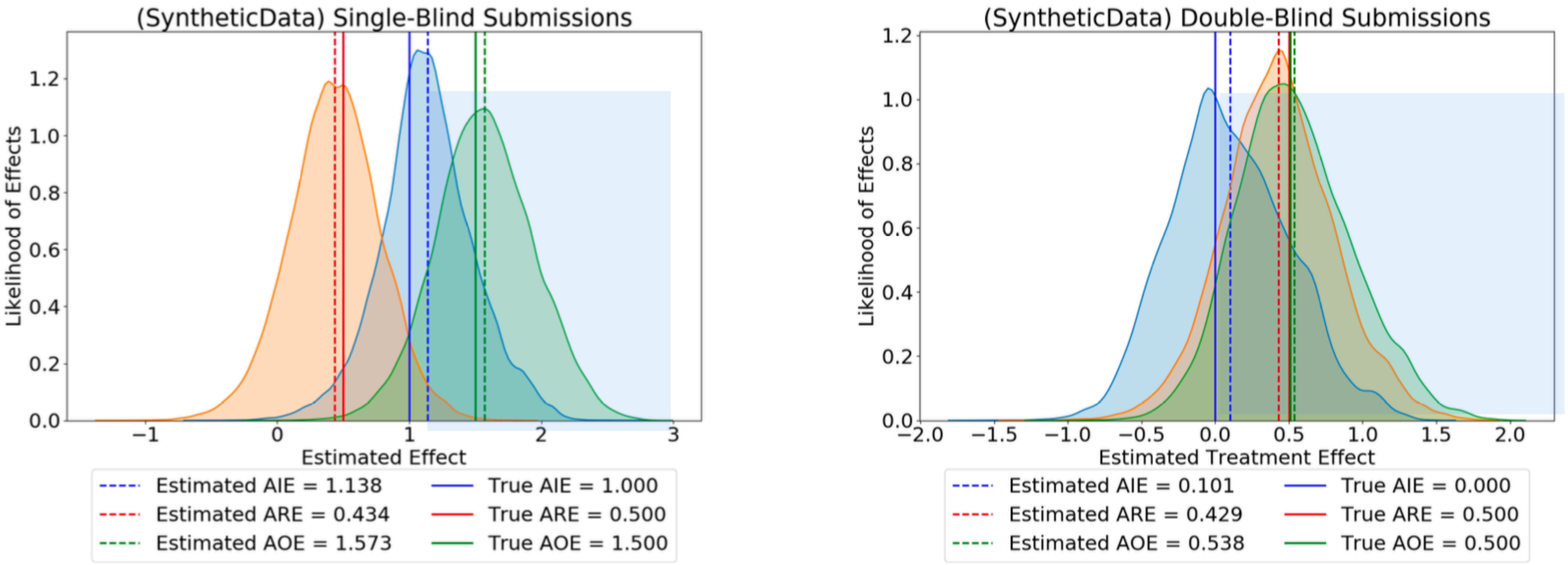
Evaluate on what?

- MIMIC-III
  - Real world ICU parameters of 59K patients
- Nationwide Inpatient Sample (NIS)
  - Real world hospital dataset
- Review Data (ReviewData)
  - Conference Submissions
- Synthetic Review Data
  - For accuracy testing

1. End to End Performance

Dataset	Tables [#]	Att. [#]	Rows [#]	Unit Table Cons.	Query Ans.
MIMIC-III	26	324	400M	6h	4.5h
NIS	4	280	8M	4m	30s
REVIEWDATA	3	7	6K	10.6s	1.2s
SYNTHETIC REVIEWDATA	3	7	300K	17.2s	1.3s

2. Correctness



3. How does Embedding affect results?

Tested On Synthetic Data