Discovering Causal Relations in Semantically-Annotated Probabilistic Business Process Diagrams

Héctor G. Ceballos, Francisco J. Cantú

Tecnológico de Monterrey

GCAI-2018: 4th Global Conference on Artificial Intelligence

Luxembourg City, Luxembourg, September 18-21, 2018

Agenda

- Business Process Diagrams
- Probabilistic BPMN normal form
- Semantic Descriptors
- A probabilistic and semantic process instance
- Causal Learning
- Decision Making
- Conclusions

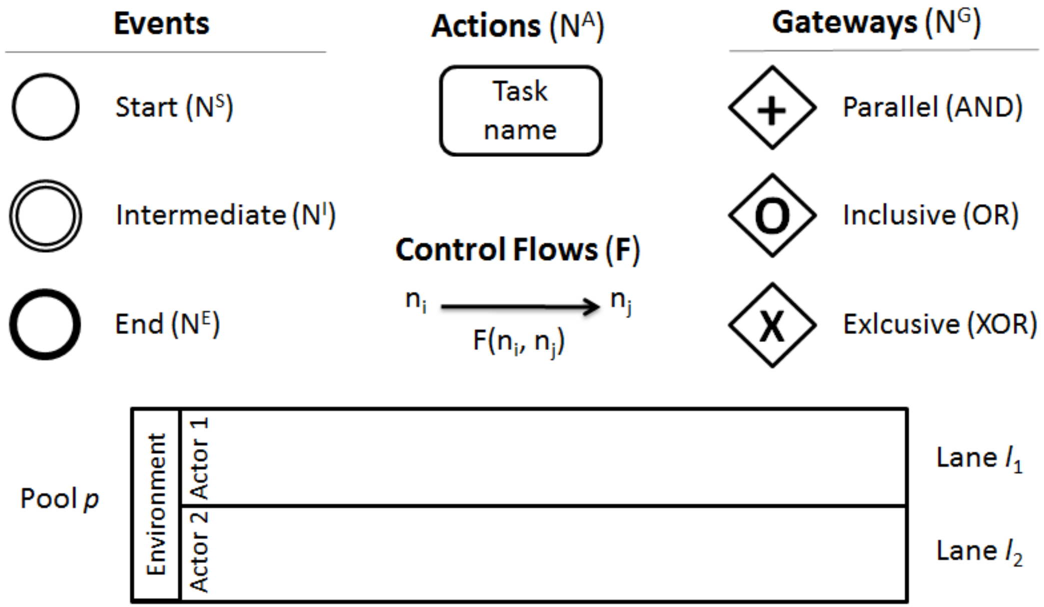
Tecnológico de Monterrey

Business Process Diagrams

- Business Process Modeling and Notation (BPMN)
  - Alternative sequencences of tasks and events.
- Stochastic BPMN
  - Based on Continuous Time-Markov Chains
  - Probability of observing a given event or task at time t.
- Probabilistic BPMN normal-form
  - Based on Bayesian Networks
  - How likely is to observe an event or a task B given that another one A occurs
- The goal: discovering causal dependencies between non-consecutive events/tasks.

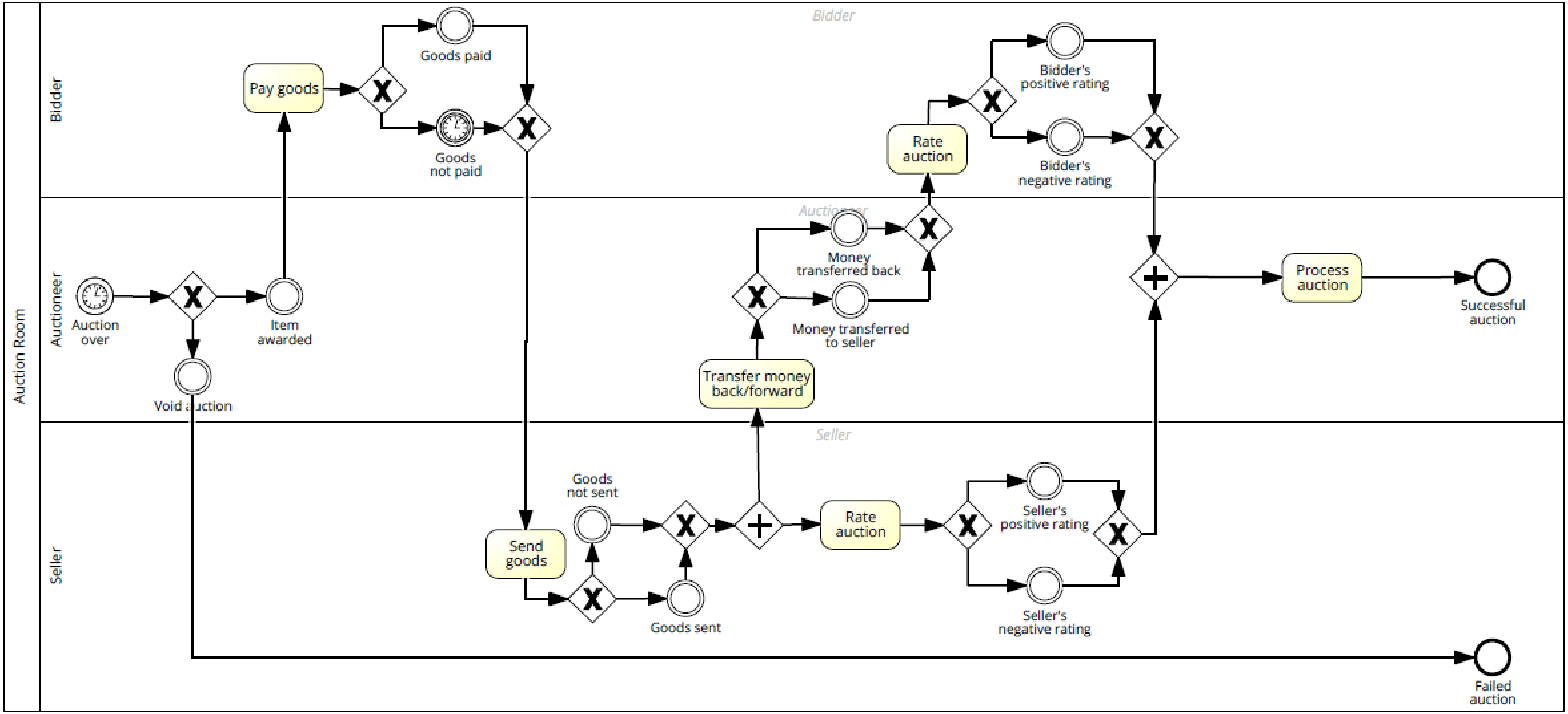
The BPMN normal form

- Business Process Diagram: **W = {P, L, N, F}**
- Lanes represent agent roles.
- DAG (G_N)



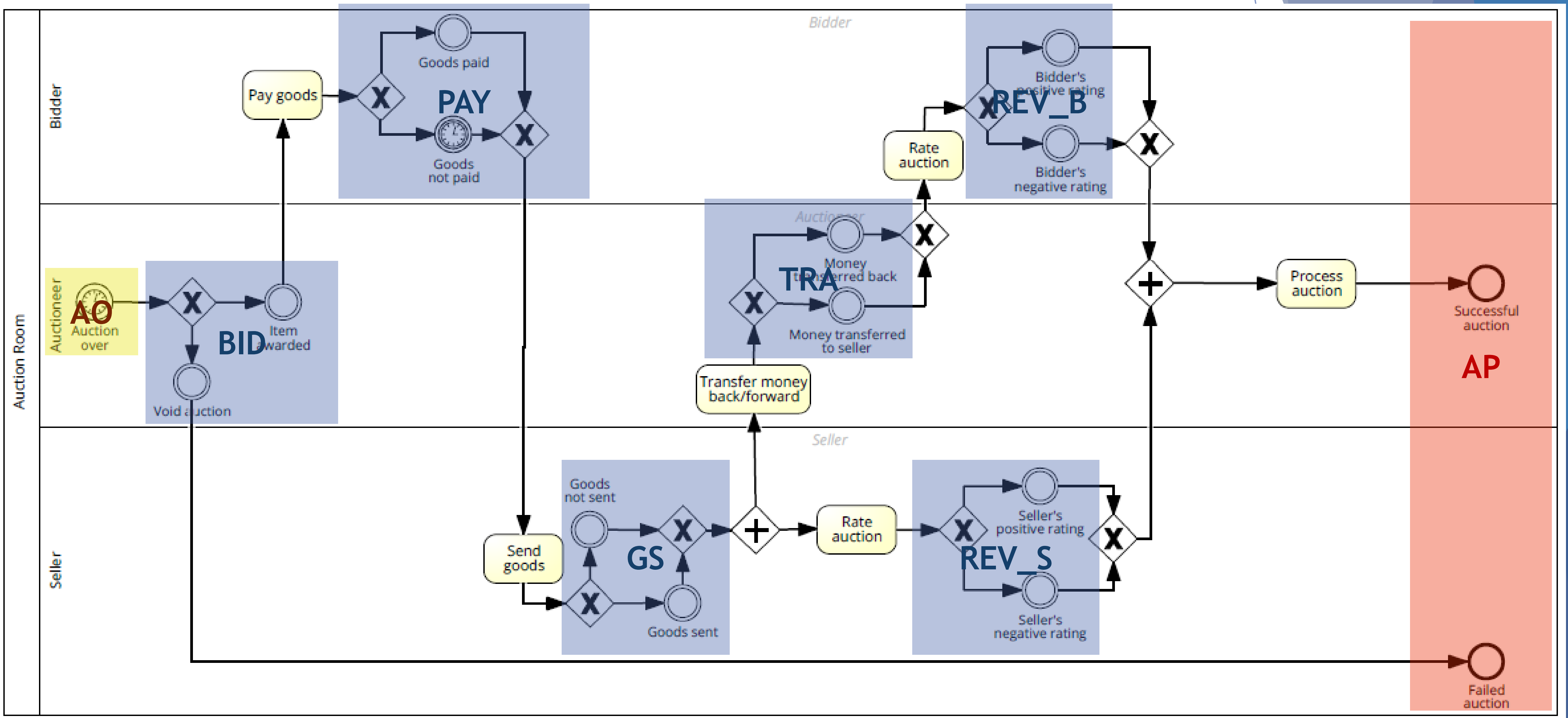
Structure		Mappings	In G'_N
Trigger	$s \rightarrow n_i$	map(s, Z_s=True) (1)	s
Outcome	$n_i \rightarrow e$	map(e_i, Z_E=e_i) (2)	e_1 (7d)
Event	$n_i \rightarrow i \rightarrow n_j$	map(i, Z_i=True) (3)	i
Actions	$n_i \rightarrow a \rightarrow n_j$	map(a, X_a=True) (6)	a
Split gateways: Decision	$n_i \rightarrow g \rightarrow n_{j-1}, n_{j-2}$	map(i, Z_g=i) (4)	g (7a)
Split gateways: Alternative	$n_i \rightarrow g \rightarrow n_{j-1}, n_{j-2}$	n_i, n_j (see above)	n_i, n_j (7c)
Merge gateways	$n_{i-1}, n_{i-2} \rightarrow g \rightarrow n_j$	n_i, n_j (see above)	n_i, n_j (7b)

Processing an electronic auction (BPD)

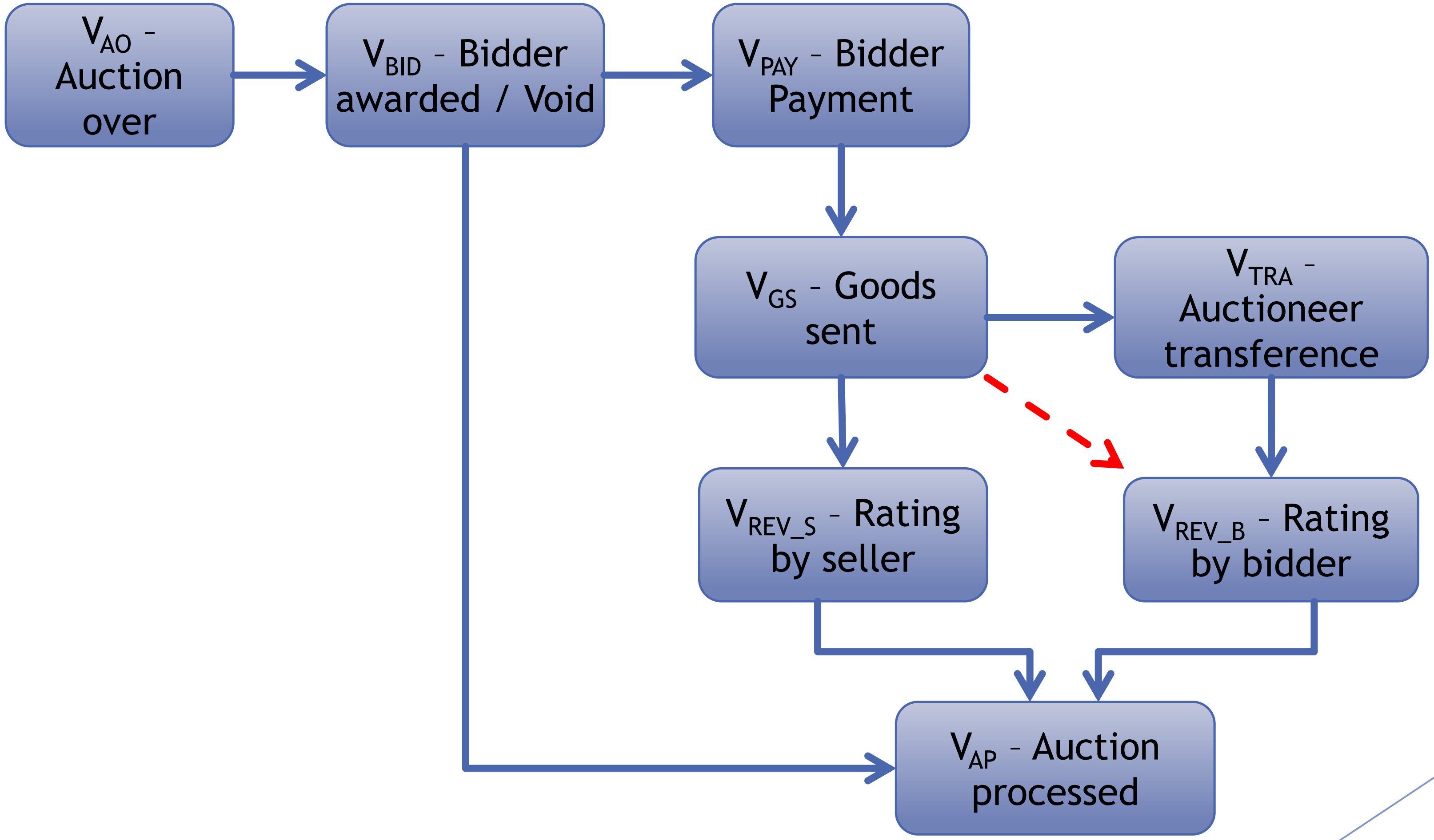


Tecnológico de Monterrey

Processing an electronic auction (BPD)



Processing an electronic auction (BN)



Tecnológico de Monterrey

Semantic Descriptors

- A **semantic descriptor** $Ann(n, Q)$ is used for describing the meaning of lanes, events and tasks, where Q is a DL conjunctive query
  - $Q = (s_1, \ldots, s_n).\{T_1, \ldots, T_m\}$

T_i { concept clauses (s rdf:type C)
relation clauses (s r s')

- A **bridge descriptor** is a semantic descriptor $Ann(n, Q_B)$ such that its conjunctive query Q_B has a single distinguished variable $Dis(Q_B) = \{s_B\}$

No.	V_i	v_{ij}	Q_{ijl}
1	V_{BID}	?bidder	$Q_{BID}() = (?bidder).\{$?a rdf : type : Auction . ?a : awardedBidder ?bidder $\}$

Tecnológico de Monterrey

A probabilistic and semantic process instance

- An instance of a process modeled through an annotated BPD W_D and compliant with a probabilistic normal form is represented by a tuple $I_W = \langle \pi_i, \overline{v_i} \rangle$.

- Example of a void auction: $I_{auction} = \langle \pi_{void}, \overline{v_{void}} \rangle$

$$\pi_{void} = \{?a : \langle things/auction\_003 \rangle, ?seller : \langle people/john \rangle\},$$

$$\bar{v}_{void} = \{V_{AO} = TRUE, V_{BID} = FALSE, V_{AP} = TRUE, V_{PAY} = FALSE, V_{GS} = FALSE\}.$$

Tecnológico de Monterrey

Structural Learning

- Structural learning on 1,000 instances based on a probabilistic distribution: 10 folds cross-validation (80% - 20%)

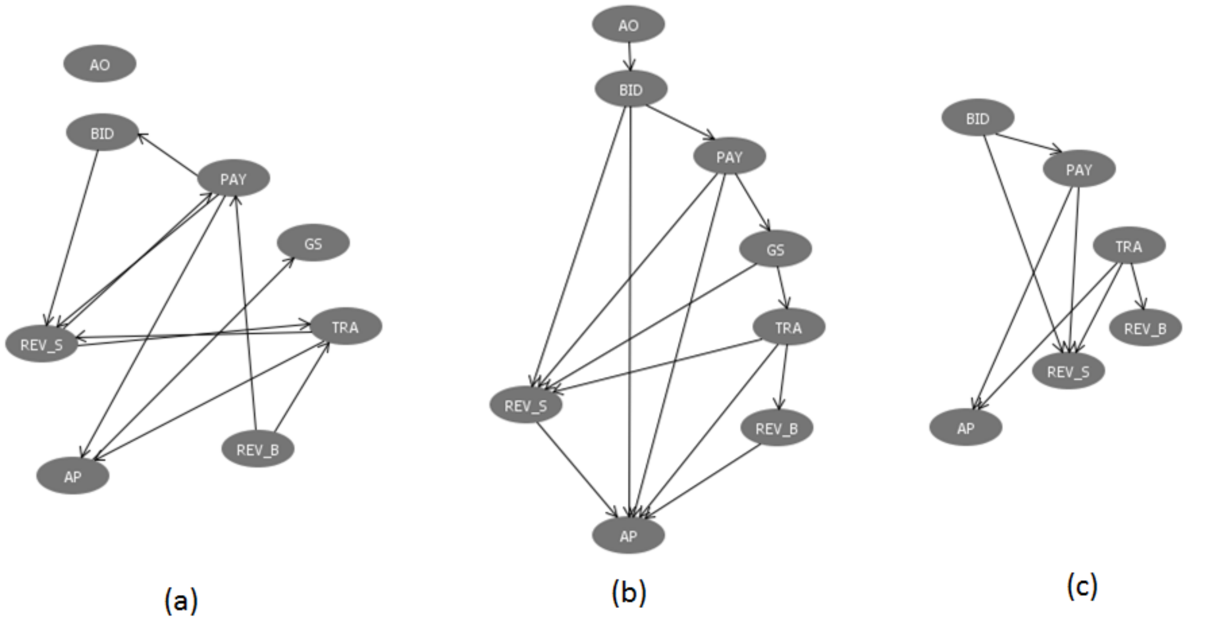
Cases	AO	BID	PAY	GS	TRA	REV_B	REV_S	AP	Comment
5%	TRUE	FALSE	FALSE	FALSE				FAILURE	Void auction
4%	TRUE	?bidder	FALSE	FALSE				FAILURE	Bidder doesn't pay item
4%	TRUE	?bidder	FALSE	FALSE			NEG	FAILURE	Bidder doesn't pay item. Seller rates negatively.
5%	TRUE	?bidder	TRUE	FALSE	BIDDER	NEG		FAILURE	Bidder pays, Seller doesn't send the item. Bidder rates negatively.
65%	TRUE	?bidder	TRUE	TRUE	SELLER	POS	POS	SUCCESS	Item sold. Rating POS-POS.
8%	TRUE	?bidder	TRUE	TRUE	SELLER	NEG	POS	SUCCESS	Item sold. Rating NEG-POS.
6%	TRUE	?bidder	TRUE	TRUE	SELLER	POS	NEG	SUCCESS	Item sold. Rating POS-NEG.
3%	TRUE	?bidder	TRUE	TRUE	SELLER	NEG	NEG	SUCCESS	Item sold. Rating NEG-NEG.

- 87% of the instances classified correctly: causal relations are not encoded in temporal precedence.

Tecnológico de Monterrey

Causal Learning

- a) Inferred Causation (IC*) discovered 5 new valid arcs and 2 new undirected arcs.
- b) Temporal precedence was used for directing these 2 arcs.
- c) A compact model with only the 7 arcs discovered by IC*.
- 100% instances classified correctly with b) and c)



Tecnológico de Monterrey

Decision Making

- The variable ?bidder in the bridge descriptor is used for selecting instances associated to a specific bidder.
- A *bad bidder* can be discovered by comparing his probabilistic distribution (<people/tom>) against the overall behavior (?bidder).

	Enriched Model		Compact Model	
A posteriori probability	?bidder	$\langle people/tom \rangle$	?bidder	$\langle people/tom \rangle$
$P(AP = SUCCESS \mid BID = ?bidder)$	0.9142	0.3924	0.8715	0.3912
$P(REV_S = POS \mid BID = ?bidder)$	0.8431	0.3091	0.8614	0.2939
$P(PAY = TRUE \mid BID = ?bidder)$	0.9153	0.3888	0.9153	0.3888

Tecnológico de Monterrey

Conclusions

- IC* discovered new causal relationships that improved process development prediction.
- Temporal precedence was used for redirecting undirected arcs.

- **Future Work**
  - Normalization of BPDs by supporting cycles.
  - Using semantic annotations for generating traces from Linked Data.

Thank you for your attention

Hector G. Ceballos (ceballos@itesm.mx)