# Expressivity Analysis for PL-Languages

*Manfred Jaeger*
Aalborg University

*Kristian Kersing, Luc De Raedt*
Freiburg University

# Semantics-based Expressivity Analysis

**The Problem**

"Alphabet soup" (L.Getoor): Prism, SLP, RBN, PRM, BLP, MLN, Blog, . . .

Questions:

- Where are these languages similar?
- Where are these languages different?
- What are the particular strengths/weaknesses of language XYZ?

First issue to investigate:

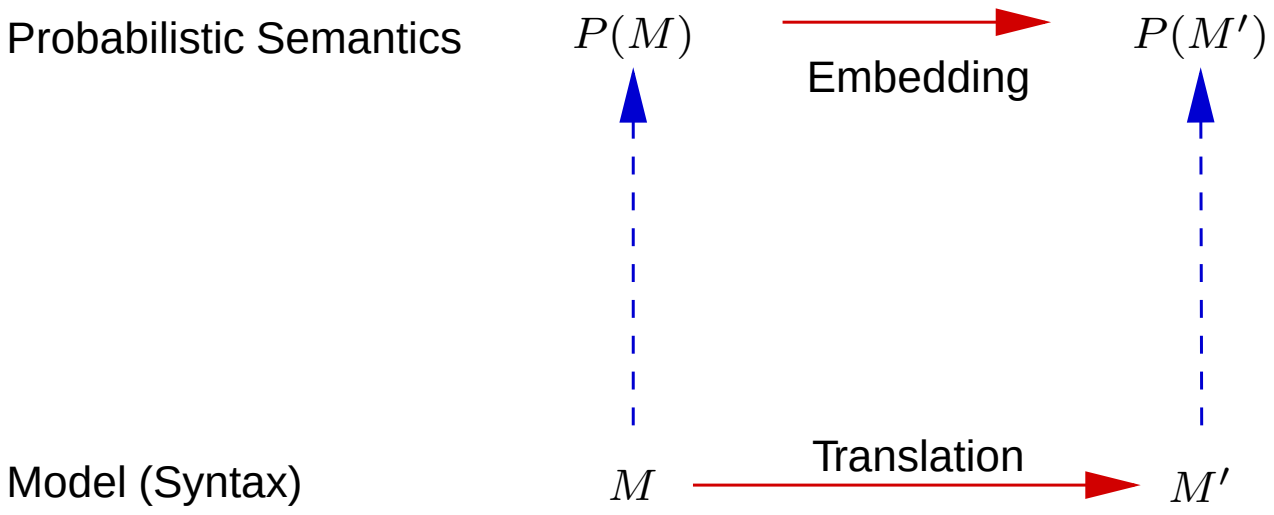
- What is the expressive power of the different languages?

Later:

- What is the complexity of inference?
- What is the complexity of learning?

## Elements of a Solution

- Goal: establish general framework with re-usable components for expressivity analysis
- Find common semantic ground
- Consider *translations* of (syntactic) models and *embeddings* of their semantics.
- A language $L'$ is *at least as expressive as* a language $L$, if each $L$-model $M$ can be translated into an $L'$-model $M'$, so that the semantics of $M'$ "contains" the semantics of $M$.



## Common Semantic Ground: Multi-valued Herbrand Interpretations

PL-languages define distributions for random variables that can be written as ground atoms:

| | | |
|---|---|---|
| *blood_pressure(tom)* | *sister(susan,tom)* | *genotype(mother(paul))* |
| *blood_pressure(susan)* | *sister(susan,paul)* | *genotype(father(paul))* |
| ... | ... | ... |

With each relation symbol is associated a (finite) state space:

$$\textit{states(blood\_pressure)}=\{\textit{high},\textit{normal},\textit{low}\}$$
$$\textit{states(sister)}=\{\textit{true},\textit{false}\}$$
$$\textit{states(genotype)}=\{\textit{AA},\textit{Aa},\textit{aa}\}$$

*Herbrand Interpretation:* assignment of a truth value to all ground atoms constructible from a vocabulary $S$ of relation, function, and constant symbols.

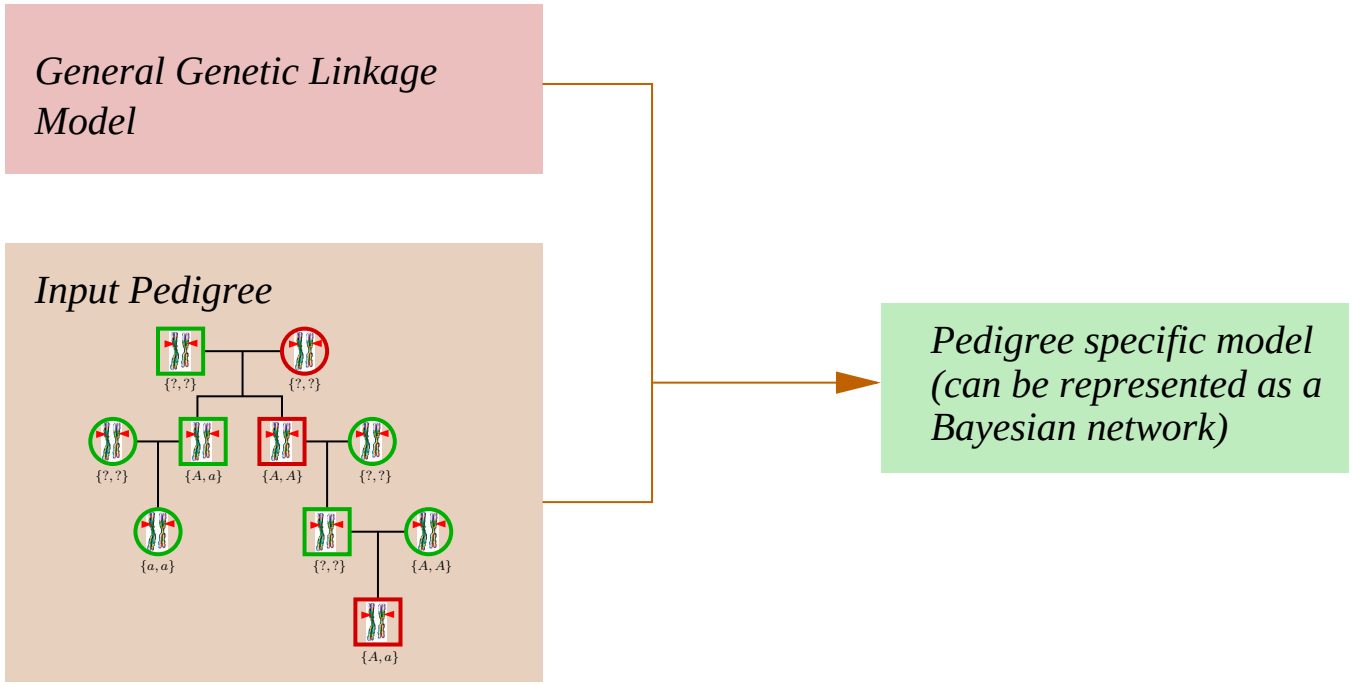
*Multi-valued Herbrand Interpretation:* assignment of a state to all ground atoms constructible from a vocabulary $S$ of relation, function, and constant symbols.

*PL-model:* defines a probability distribution over all Multi-valued Herbrand Interpretations for a given vocabulary $S$.

Any PL-model can be represented by an ordinary Bayesian network. Are PL-languages just shorthand notations for large Bayesian networks?

# Modularity of Representations

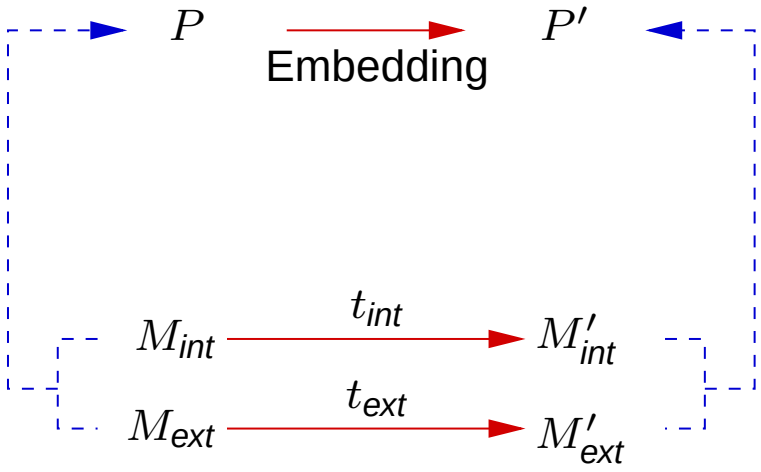
The power and usefulness of PL-languages derives from the fact that they split the specification of a complex model into a generic (*intensional*) and a domain-specific (*extensional*) part:



A (preliminary) analysis of several languages:

| | Intensional | Extensional |
|---|---|---|
| RBN | rbn | Input Structure |
| PRM | prm | Skeleton Structure |
| BLP | intensional part | extensional part |
| MLN | mln | constants |
| Prism | ground atoms with msw's in SLD tree | ground atoms without msw's in SLD tree |

Updated plan:



# Formalization

**Embeddings**

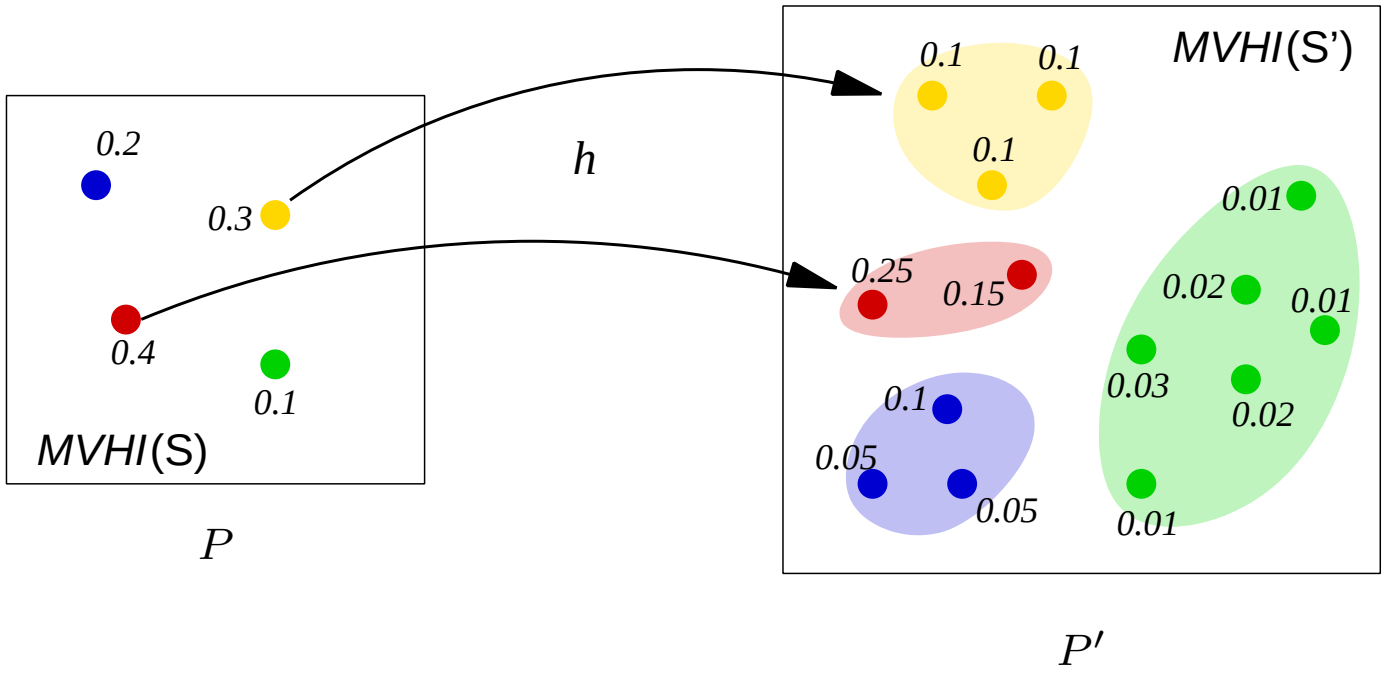
$P$: probability distributions over *MVHI*(S)
$P'$: probability distributions over *MVHI*(S')
An embedding of $P$ in $P'$ is a mapping

$$h : \mathit{MVHI}(S) \mapsto 2^{\mathit{MVHI}(S')}$$

such that for all $w, w' \in \mathit{MVHI}(S)$:

$$P(w) = P'(h(w)) \text{ and } h(w) \cap h(w') = \emptyset$$

Write $P \preceq P'$ if there is such an embedding.




$P$ $P'$

If $P \preceq P'$, then every probabilistic query about $P$ can be answered from the model $P'$ (one can consider weaker forms of embeddings, so that only restricted types of queries for $P$ are supported by $P'$).

## Putting Everything Together...

Language $L'$ is at least as expressive as $L$, $L \preceq L'$, if

$$\exists t_{int} \forall M_{int} \exists t_{ext} \forall M_{ext} \quad P(M_{int}, M_{ext}) \preceq P(t_{int}(M_{int}), t_{ext}(M_{ext}))$$

## Example Result

$$MLN \preceq RBN \quad \text{(precisely: } MLN \preceq_c RBN)$$