

A Complete Characterization of Projectivity for Statistical Relational Models

Manfred Jaeger* and Oliver Schulte**

* Aalborg University

** Simon Fraser University

Possible Worlds

Heterogeneous relational structures that allow multiple types of relationships and nodes:



Possible Worlds

Heterogeneous relational structures that allow multiple types of relationships and nodes:



Generative probabilistic model

We are concerned with **families of distributions**:

$$P^{(n)} \quad (n \in \mathbb{N})$$

where $P^{(n)}$ is a distribution over n -worlds.



Exchangeability

We always assume that all distributions are **exchangeable**:

$$P^{(3)}\left(\begin{array}{c} a \text{ (red)} \\ \quad \quad \quad b \text{ (blue)} \\ \quad \quad \quad | \\ \quad \quad \quad c \text{ (red)} \end{array} \right) = P^{(3)}\left(\begin{array}{c} \quad \quad \quad n_{107} \text{ (red)} \\ \quad \quad \quad / \quad \backslash \\ n_{312} \text{ (red)} \quad n_{87} \text{ (blue)} \end{array} \right)$$

Isomorphic worlds have the same probability (names of nodes do not matter)!

Exchangeability

We always assume that all distributions are **exchangeable**:

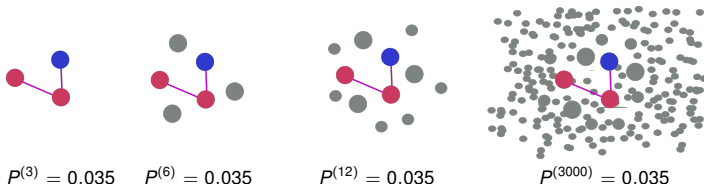
$$P^{(3)}(\begin{array}{c} a \quad b \\ \quad \diagup \quad \diagdown \\ \quad c \end{array}) = P^{(3)}(\begin{array}{c} n_{107} \\ \diagup \quad \diagdown \\ n_{312} \quad n_{87} \end{array})$$

Isomorphic worlds have the same probability (names of nodes do not matter)!

Why families of generative models?

- ▶ inherently **multi-relational**.
- ▶ **inductive**: learning from *training worlds* $\omega_1, \dots, \omega_N$ (of different sizes); prediction for *query entities* embedded in a new world ω_q .
- ▶ **flexible inference**: computation of arbitrary conditional probabilities.

Projectivity



The probability of a fixed substructure does not depend on the size of the domain it is embedded in.

Can be desirable, because:

- ▶ Exact size of domain may be unknown or changing over time.
- ▶ Inference complexity independent of domain size (inference is *extremely lifted!*)
- ▶ Avoid *degeneracy*: extreme probabilities (0 or 1) for queries as $n \rightarrow \infty$

What we already know

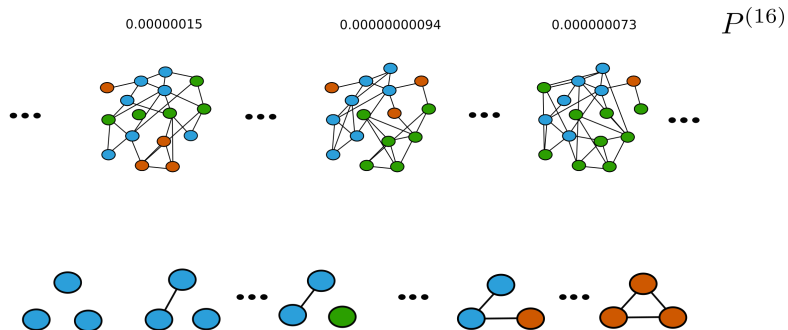
- ▶ Most SRL frameworks are not projective.
- ▶ Some syntactic fragments of SRL modeling languages identified as projective [Jaeger, Schulte, StarAI 2018].

Our question

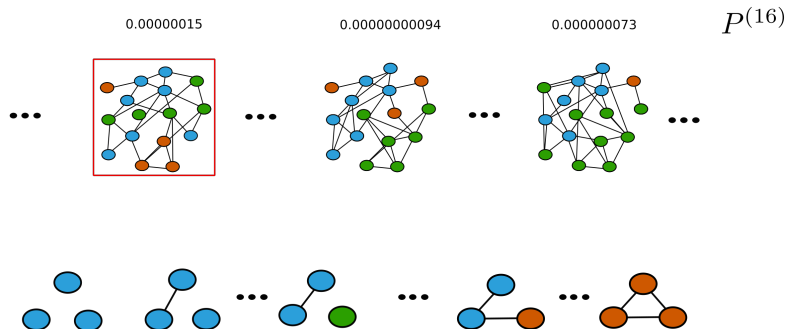
- ▶ What exactly are projective models?
- ▶ Can we identify a representation language that can represent **all** projective models, and **only** projective models?

➡ On our way towards the answer to these questions, we clarify relationships with several other structural properties of generative families $(P^{(n)})_{n \in \mathbb{N}}$.

Sampling and Extending

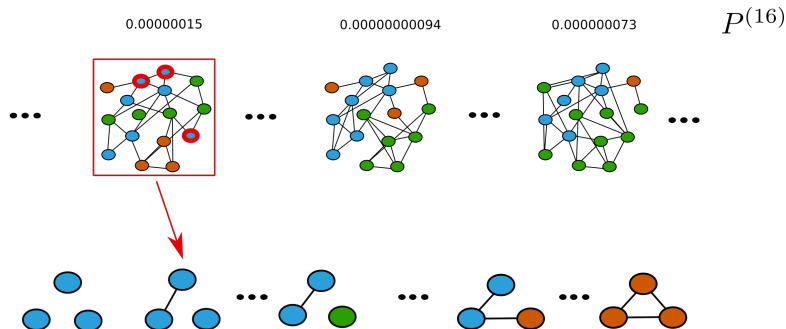


Given: $P^{(n)}$



Given: $P^{(n)}$

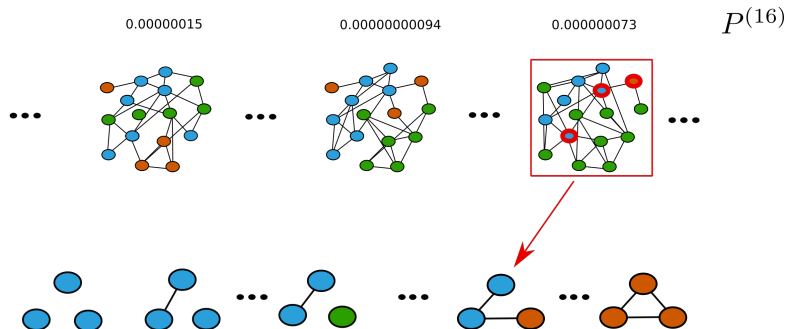
...pick n -world according to $P^{(n)}$



Given: $P^{(n)}$

... pick n -world according to $P^{(n)}$

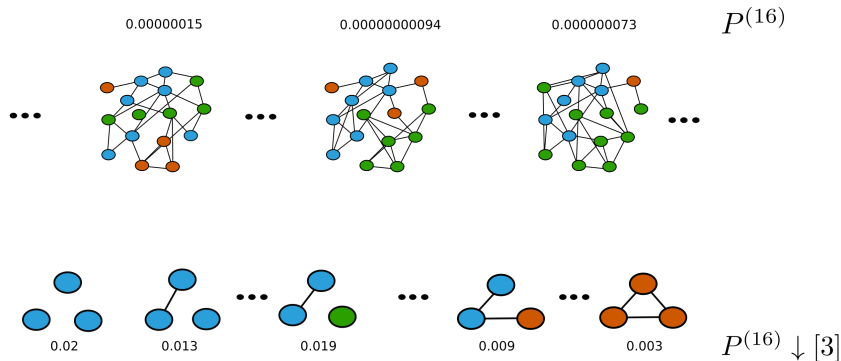
... select ordered m -tuple of elements \rightsquigarrow sampled m -world



Given: $P^{(n)}$

... pick n -world according to $P^{(n)}$

... select ordered m -tuple of elements \rightsquigarrow sampled m -world



Given: $P^{(n)}$

... pick n -world according to $P^{(n)}$

... select ordered m -tuple of elements \rightsquigarrow sampled m -world

➡ Obtain sampling/marginal distribution $P^{(n)} \downarrow [m]$

Projectivity: precise

The family $(P^{(n)})_n$ is **projective**, if for all $m < n$:

$$P^{(n)} \downarrow [m] = P^{(m)}.$$

Extendability

$P^{(m)}$ is (projective) **extendable** if for every n there exists $P^{(n)}$ such that

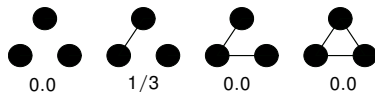
$$P^{(n)} \downarrow [m] = P^{(m)}$$

and the $P^{(n)}$ form a projective family.

Why care about extendability?

- ▶ Often use weights/probabilities for small sub-structures (“features”, “motifs”) as the basis for defining a model:
 - ▶ *Exponential random graph models, Markov logic networks*
 - ▶ Halpern, Bacchus, Grove, Koller: *random worlds approach*
 - ▶ Kuželka, Wang, Davis, Schockaert (AAAI 2018): *relational marginals*
- ▶ A distribution $P^{(m)}$ (small m) is a prototype for such a specification
- ▶ Question then: given a specific $P^{(m)}$, can it be consistently generalized to arbitrary $n > m$?

Counterexample



is not extendable (not the marginal of any $P^{(n)}$ with $n \geq 5$).

[related example: Kuželka et al., AAAI 18]

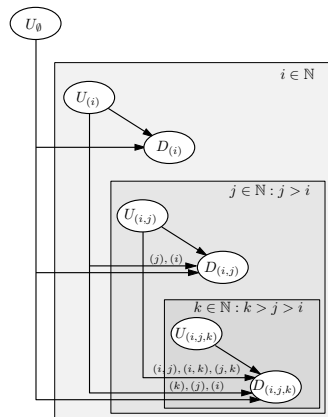
A first equivalence

$$P^{(m)} \text{ extendable} \Leftrightarrow P^{(m)} \text{ projective extendable}$$

AHK Representations and Representation Theorem

Adaptation of representation of infinite exchangeable arrays by Aldous (1981), Hoover (1979), Kallenberg (2006).

Plate representation for a given *relational signature* S with relations of *maximal arity* a .



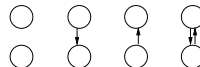
Model components:

D -variables: encode possible world by collecting for $m = 1, \dots, a$ all the m -arity data.

Example $S = \{\text{red}/1, \text{edge}/2\}$. State space for $D_{(i)}$ variables:



State space for $D_{(i,j)}$ variables:



U -variables: i.i.d., uniform on $[0, 1]$

Conditional distribution of D 's given U 's defined by **deterministic functions** $f^{(m)}$ $m = 1, \dots, a$.

We obtain characterizations of projectivity:

... for a given $P^{(m)}$:

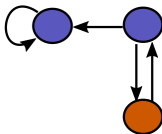
$$P^{(m)} \text{ extendable} \Leftrightarrow P^{(m)} \text{ projective extendable} \Leftrightarrow P^{(m)} \text{ has AHK representation}$$

... for a family $(P^{(n)})_n$:

$$(P^{(n)})_n \text{ projective} \Leftrightarrow (P^{(n)})_n \text{ has AHK representation}$$

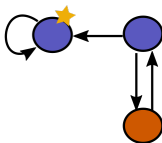
➡ AHK models are a maximally expressive framework for “very well behaved” families

The AHK characterization leads to insights into the structure of projective families:



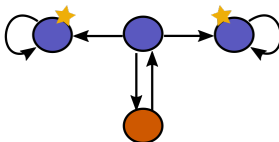
Take a world with positive probability according to $P^{(n)}$,

The AHK characterization leads to insights into the structure of projective families:



Take a world with positive probability according to $P^{(n)}$,
... select an arbitrary entity

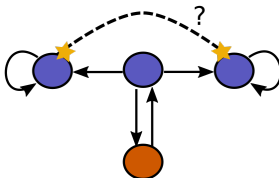
The AHK characterization leads to insights into the structure of projective families:



Take a world with positive probability according to $P^{(n)}$,

- ... select an arbitrary entity
- ... create a clone of that entity

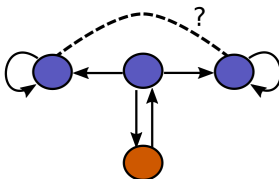
The AHK characterization leads to insights into the structure of projective families:



Take a world with positive probability according to $P^{(n)}$,

- ... select an arbitrary entity
- ... create a clone of that entity
- ... relations between clone and original are unspecified

The AHK characterization leads to insights into the structure of projective families:



Take a world with positive probability according to $P^{(n)}$,

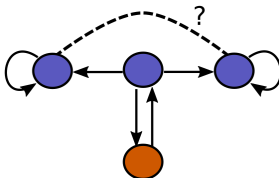
- ... select an arbitrary entity

- ... create a clone of that entity

- ... relations between clone and original are unspecified

➡ The resulting *set of* $n + 1$ -worlds then has a nonzero probability in $P^{(n+1)}$

The AHK characterization leads to insights into the structure of projective families:



Take a world with positive probability according to $P^{(n)}$,

... select an arbitrary entity

... create a clone of that entity

... relations between clone and original are unspecified

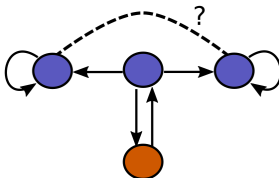
➡ The resulting set of $n + 1$ -worlds then has a nonzero probability in $P^{(n+1)}$

Counterexample revisited

If $(P^{(n)})_n$ projective:

$$P^{(3)}(\bullet \overset{\bullet}{\nearrow} \bullet) > 0$$

The AHK characterization leads to insights into the structure of projective families:



Take a world with positive probability according to $P^{(n)}$,

... select an arbitrary entity

... create a clone of that entity

... relations between clone and original are unspecified

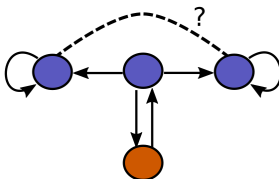
➡ The resulting *set of* $n + 1$ -worlds then has a nonzero probability in $P^{(n+1)}$

Counterexample revisited

If $(P^{(n)})_n$ projective:

$$P^{(3)}(\bullet \overset{\bullet}{\nearrow} \bullet) > 0 \Rightarrow P^{(2)}(\bullet \overset{\bullet}{\nearrow}) > 0$$

The AHK characterization leads to insights into the structure of projective families:



Take a world with positive probability according to $P^{(n)}$,

... select an arbitrary entity

... create a clone of that entity

... relations between clone and original are unspecified

➡ The resulting *set of* $n + 1$ -worlds then has a nonzero probability in $P^{(n+1)}$

Counterexample revisited

If $(P^{(n)})_n$ projective:

$$P^{(3)}(\text{graph 1}) > 0 \Rightarrow P^{(2)}(\text{graph 2}) > 0 \Rightarrow P^{(3)}(\text{graph 3}) > 0$$

Done:

- ▶ Comprehensive characterization of projectivity and related properties

Todo:

- ▶ Fuller understanding of tradeoff for projective models:
 - ▶ Benefits: computational, statistical tractability. Semantic robustness.
 - ▶ Drawbacks: limits of expressivity
- ▶ Practical implementation of AHK models:
 - ▶ how to represent and learn the deterministic functions that map U -variables to D -variables?