On Projectivity in Markov Logic Networks

Sagar Malhotra[1,2] **and Luciano Serafini**[1]
[1]Fondazione Bruno Kessler, Italy
[2]University of Trento, Italy
smalhotra@fbk.eu, serafini@fbk.eu

Abstract

Markov Logic Networks (MLNs) define a probability distribution on relational structures over varying domain sizes. Like most relational models, MLNs do not admit consistent marginal inference over varying domain sizes i.e. marginal probabilities depend on the domain size. Furthermore, MLNs learned on a fixed domain do not generalize to domains of different sizes. In recent works, connections have emerged between domain size dependence, lifted inference, and learning from a sub-sampled domain. The central idea of these works is the notion of *projectivity*. The probability distributions ascribed by projective models render the marginal probabilities of sub-structures independent of the domain cardinality. Hence, projective models admit efficient marginal inference. Furthermore, projective models potentially allow efficient and consistent parameter learning from sub-sampled domains. In this paper, we characterize the necessary and sufficient conditions for a two-variable MLN to be projective. We then isolate a special class of models, namely Relational Block Models (RBMs). In terms of data likelihood, RBMs allow us to learn the best possible projective MLN in the two-variable fragment. Furthermore, RBMs also admit consistent parameter learning over sub-sampled domains.

Introduction

Statistical Relational Learning (Getoor and Taskar 2007; Raedt et al. 2016) (SRL) is concerned with representing and learning probabilistic models over relational structures. Many works have observed that SRL frameworks exhibit unwanted behaviors over varying domain sizes (Poole et al. 2014; Mittal et al. 2019). These behaviors make models learned from a fixed or a sub-sampled domain unreliable for inference over larger (or smaller) domains (Mittal et al. 2019). Drawing on the works of Shalizi and Rinaldo (Shalizi and Rinaldo 2013) on Exponential Random Graphs (ERGMs), Jaeger and Schulte (Jaeger and Schulte 2018) have recently introduced the notion of *projectivity* as a strong form of guarantee for good scaling behavior in SRL models. A projective model requires that the probability of any given query, over arbitrary m domain objects, is completely independent of the domain size.

Jaeger and Schulte (Jaeger and Schulte 2018) identify restrictive fragments of SRL models to be projective. But whether these fragments are complete characterization of projectivity, remains an open problem.

In this paper, our goal is to characterize projectivity for a specific class of SRL models, namely Markov Logic Networks (MLNs) (Richardson and Domingos 2006). MLNs are amongst the most prominent template-based SRL models. An MLN is a Markov Random Field with features defined in terms of function-free weighted First Order Logic (FOL) formulae. Jaeger and Schutle (Jaeger and Schulte 2018) show that an MLN is projective if - any pair of atoms in each of its formulae share the same set of variables. We show that this characterization is not complete. Furthermore, we completely characterize projectivity for the class of MLNs with at most 2 variables in their formulae. Our charecterization leads to a parametric restriction that can be easily incorporated into any MLN learning algorithm. We also identify a special class of projective models, namely the Relational Block Models (RBMs). Any projective MLN in the two variable fragment can be expressed as an RBM. We show that the training data likelihood due to the maximum likelihood RBM is greater than or equal to the training data likelihood due to any other projective MLN in the two variable fragment. RBMs also admit consistent maximum likelihood estimation. Hence, RBMs are projective models that admit consistent and efficient learning from sub-sampled domains.

The paper is organized as follows: We first contextualize our work w.r.t the related works in this domain. We then provide some background and notation on FOL and relational structures. We also elaborate on the fragment of FOL with at most two variables i.e. FO^2 and define the notion of FO^2 interpretations as multi-relational graphs. We also overview some results on Weighted First Order Model Counting. In the subsequent section, we provide a parametric representation for any MLN in the two variable fragment. We then dedicate a section to the main result of this paper i.e. the necessary and sufficient conditions for an MLN in the two variable fragment to be projective. Based on the projectivity criterions we identify a special class of models namely Relational Block Models. We dedicate a complete section to RBMs and elaborate on their useful properties. We then move on to a formal comparison between the previous characterizations and the presented characterization of projectivity in MLNs. Finally, we discuss the consistency and efficiency aspects of learning for projective MLNs and RBMs.

Related Work

Projectivity has emerged as a formal notion of interest through multiple independent lines of works across ERGM and SRL literature. The key focus of these works have been analyzing (Snijders 2010; Poole et al. 2014) or mitigating (Jain, Barthels, and Beetz 2010; Mittal et al. 2019) the effects of varying domain sizes on relational models. The major step in formalizing the notion of projectivity can be attributed to Shalizi and Rinaldo (Shalizi and Rinaldo 2013). The authors both formalize and characterize the sufficient and necessary conditions for ERGMs to be projective. It is interesting to note that their projectivity criterion is strictly structural i.e. they put no restrictions on parameter values but rather inhibit the class of features that can be defined as sufficient statistics in ERGMs. In contrast our results w.r.t MLN are strictly parametric (which may correspond to non-trivial structural restrictions as well). With respect to SRL, the notion of projectivity was first formalized by Jaeger and Schulte (Jaeger and Schulte 2018), they show some restrictive fragments of SRL models to be projective. Jaeger and Schulte (Jaeger and Schulte 2020) significantly extend the scope of projective models by characterizing necessary and sufficient conditions for an arbitrary model on relational structures to be projective. Their characterization is expressed in terms of the so called AHK models. But as they conclude in (Jaeger and Schulte 2020), expressing AHK models in existing SRL frameworks remains a significant open challenge. Hence, a complete characterization of projectivity in most SRL languages is still an open problem. Weitkamper (Weitkämper 2021) has shown that the characterization of projectivity provided by Jaeger and Schulte (Jaeger and Schulte 2018), for probabilistic logic programs under distribution semantics, is indeed complete. In this work, we will extend this characterization to the two variable fragment of Markov Logic Networks.

Another correlated problem to projectivity is learning from sub-sampled or smaller domains. In the relational setting projectivity is not a sufficient condition for consistent learning from sub-sampled domains (Jaeger and Schulte 2018). Mittal et. al. have proposed a solution to this problem by introducing domain-size dependent scale-down factors (Mittal et al. 2019) for MLN weights. Although empirically effective, the scale-down factors are not known to be a statistically sound solution. On the other hand, Kuzelka et. al. (Kuzelka et al. 2017), provide a statistically sound approach to approximately obtain the correct distribution for a larger domain. But their approach requires estimating the relational marginal polytope for the larger domain and hence, offers no computational gains w.r.t learning from a sub-sampled domain. In this work, we will provide a statistically sound approach for efficiently estimating a special class of projective models (namely, RBM) from sub-sampled domains. We also show that our approach provides consistent parameter estimates in an efficient manner and is better than estimating any projective MLN in the two variable fragment (in terms of data likelihood maximisation).

Background

Basic Definitions. We use the following basic notation. The set of integers $\{1, ..., n\}$ is denoted by $[n]$. We use $[m : n]$ to denote the set of integers $\{m, ..., n\}$. Wherever the set of integers $[n]$ is obvious from the context we will use $[\overline{m}]$ to represent the set $[m+1 : n]$. We use $\boldsymbol{k} = \langle k_1, ..., k_m \rangle$ to denote an n-partition i.e. $k_i \in \mathbb{Z}^+$ and $\sum_{i\in[m]} k_i = n$. We will also use multinomial coefficients denoted by

$$\binom{n}{k_1, ..., k_m} = \binom{n}{\boldsymbol{k}} = \frac{n!}{\prod_{i\in[m]} k_i!}$$

First Order Logic and Relational Substructures. We assume a function-free First Order Logic (FOL) language $\mathcal{L}$ defined by a set of variables $\mathcal{V}$ and a set of relational symbols $\mathcal{R}$. We use Δ to denote a domain of n constants. For $a_1, ..., a_k \in \mathcal{V} \cup \Delta$ and $R \in \mathcal{R}$, we call $R(a_1, ...a_k)$ an *atom*. A *literal* is an atom or the negation of an atom. If $a_1, ..., a_k \in \mathcal{V}$, then the atom is called a *first order atom*, whereas if $a_1, ..., a_k \in \Delta$, then it's called a *ground atom*. We use $\mathcal{F}$ to denote the set of first order atoms and $\mathcal{G}$ to denote the set of ground atoms. A *world* or an *interpretation* $\omega : \mathcal{G} \rightarrow \{\mathbf{T}, \mathbf{F}\}$ is a function that maps each ground atom to a boolean. The set of interpretations ω, in the language $\mathcal{L}$ and the domain Δ of size n, is denoted by $\Omega^{(n)}$. We say that $\omega \in \Omega^{(n)}$, has a size n and is also called an n-world. For a subset $\mathrm{I} \subset \Delta$, we use $\omega \downarrow \mathrm{I}$ to denote the partial interpretation induced by I. Hence, $\omega \downarrow \mathrm{I}$ is an interpretation over the ground atoms containing only the domain elements in I.

Example 1. *Let us have a language with only one relational symbol R of arity 2 and a domain $\Delta = \{a, b, c\}$. Let us have the following interpretation ω:*

$R(a,a)$	$R(a,b)$	$R(a,c)$	$R(b,a)$	$R(b,b)$	$R(b,c)$	$R(c,a)$	$R(c,b)$	$R(c,c)$
T	T	F	T	T	F	T	T	F

then $\omega \downarrow \{a, b\}$ is given as:

$R(a,a)$	$R(a,b)$	$R(b,a)$	$R(b,b)$
T	**T**	**F**	**T**

For most of our purposes, we will be able to assume w.l.o.g that $\Delta = [n]$.

FO^2, m-Types and m-Tables. FO^2 is the fragment of FOL with two variables. We will use the notion of 1-types, 2-type, and 2-tables as presented in (Kuusisto and Lutz 2018). A 1-type is a conjunction of a maximally consistent set of first order literals containing only one variable. For example, in an FO^2 language on the unary predicate A and binary predicate R, $A(x) \wedge R(x,x)$ and $A(x) \wedge \neg R(x,x)$ are examples of 1-types in variable x. A 2-table is a conjunction of maximally consistent first order literals containing exactly two distinct variables. Extending the previous example, $R(x,y) \wedge \neg R(y,x)$ and $R(x,y) \wedge R(y,x)$ are instances of 2-tables. We assume an arbitrary order on the 1-types and 2-tables, hence, we use $i(x)$ to denote the i^{th} 1-type and $l(x,y)$ to denote the l^{th} 2-table. Finally, a 2-type is a conjunction of the form $i(x) \wedge j(y) \wedge l(x,y) \wedge (x \neq y)$ and we use $ijl(x,y)$ to represent it. In a given interpretation ω, we say a constant c realizes the i^{th} 1-type if $\omega \models i(c)$, we say a pair of constants (c, d) realizes the l^{th} 2-table if $\omega \models l(c, d)$ and (c, d) realizes the 2-type $ijl(x, y)$ if $\omega \models ijl(c, d)$. We call the 2-type $ijl(y, x)$ the dual of $ijl(x, y)$ and denote it by $\underline{ijl}(x, y)$. We will use u to denote the number of 1-types and b to denote the number of 2-tables in a given FO^2 language.

Interpretations as Multi-relational Graphs. Given an FO^2 language $\mathcal{L}$ with interpretations defined over the domain $\Delta = [n]$, we can represent an interpretation $\omega \in \Omega^{(n)}$ as a multi-relational graph $(\boldsymbol{x}, \boldsymbol{y})$. This is achieved by defining $\boldsymbol{x} = (x_1, ..., x_n)$ such that $x_q = i$ if $\omega \models i(q)$ and by defining $\boldsymbol{y} = (y_{12}, y_{13}, ...y_{qr}, ..., y_{n-1,n})$, where $q < r$, such that $y_{qr} = l$ if $\omega \models l(q, r)$. We also define $k_i = k_i(\boldsymbol{x}) = k_i(\omega) := |\{c \in \Delta : c \models i(c)\}|$, $h_l^{ij} = h_l^{ij}(\boldsymbol{y}) = h_l^{ij}(\omega) := |\{(c, d) \in \Delta^2 : \omega \models ijl(c, d)\}|$ and for any $D \subseteq \Delta^2$, $h_l^{ij}(D) = h_l^{ij}(\omega, D) := |\{(c, d) : \omega \models ijl(c, d) \text{ and } (c, d) \in D\}|$. Notice that $\sum_{i \leq j} \sum_{l \in [b]} h_l^{ij} = \binom{n}{2}$ and $\sum_{l \in [b]} h_l^{ij} = \boldsymbol{k}(i, j)$, where $\boldsymbol{k}(i, j)$ is defined in equation (3) . We use $(\boldsymbol{x}_{\mathrm{I}}, \boldsymbol{y}_{\mathrm{I}})$ to represent the multi-relational graph for $\omega \downarrow \mathrm{I}$. Throughout this paper we will use an interpretation ω and it's multi-relational graph $(\boldsymbol{x}, \boldsymbol{y})$ interchangeably.

Weighted First Order Model Counting in FO^2. We will briefly review Weighted First Order Model Counting (WFOMC) in FO^2 as presented in (Malhotra and Serafini 2021). WFOMC is formally defined as follows:

$$\mathrm{WFOMC}(\Phi, n) := \sum_{\omega \in \Omega^{(n)} : \omega \models \Phi} w(\omega)$$

where Φ is an FOL formula, n is the size of the domain and w is a weight function that maps each interpretation ω to a positive real. First Order Model Counting (FOMC) is the special case of WFOMC, where for all $\omega \in \Omega^{(n)}$, $w(\omega) = 1$. We assume that w does not depend on individual domain constants, which implies that w assigns same weight to two interpretations which are isomorphic under the permutation of domain elements.

A universally quantified FO^2 formula $\forall xy.\Phi(x, y)$ can be equivalently expressed as $\forall xy.\Phi(\{x, y\})$, where $\Phi(\{x, y\})$ is defined as $\Phi(x, x) \wedge \Phi(x, y) \wedge \Phi(y, x) \wedge \Phi(y, y) \wedge (x \neq y)$. A *lifted interpretation* denoted by $\tau : \mathcal{F} \rightarrow \{\mathbf{T}, \mathbf{F}\}$ assigns boolean values to first order atoms. The truth value of the quantifier free formula $\Phi(x, y)$ under a lifted interpretation τ, denoted by $\tau(\Phi(x, y))$, is computed by applying classical semantics of the propositional connectives to the truth assignments of atoms of $\Phi(\{x, y\})$ under τ. We then define

$$n_{ijl} := |\{\tau \mid \tau \models \Phi(\{x, y\}) \wedge ijl(x, y)\}| \tag{1}$$

and $n_{ij} := \sum_{l \in [b]} n_{ijl}$. First Order Model Counting for a universally quantified formula $\forall xy.\Phi(x, y)$ is then given as:

$$\mathrm{FOMC}(\forall xy.\Phi(x, y), n) = \sum_{\boldsymbol{k}} \binom{n}{\boldsymbol{k}} \prod_{\substack{i \leq j \\ i, j \in [b]}} n_{ij}^{\boldsymbol{k}(i,j)} \tag{2}$$

where $\boldsymbol{k} = \langle k_1, \ldots, k_u \rangle$ is a u-tuple of non-negative integers, $\binom{n}{\boldsymbol{k}}$ is the multinomial coefficient and

$$\boldsymbol{k}(i, j) = \begin{cases} \frac{k_i(k_i - 1)}{2} & \text{if } i = j \\ k_i k_j & \text{otherwise} \end{cases} \tag{3}$$

Intuitively, k_i represents the number of constants c of 1-type i. Also a given constant realizes exactly one 1-type. Hence, for a given $\boldsymbol{k}$, we have $\binom{n}{\boldsymbol{k}}$ possible ways of realizing k_i 1-types. Furthermore, given a pair of constants c and d such that c is of 1-type i and d is of 1-type j, the number of extensions to the binary predicates containing both c and d, such that the extensions are a model of $\forall xy.\Phi(x, y)$, is given by n_{ij} independently of all other constants. Finally, the exponent $\boldsymbol{k}(i, j)$ accounts for all possible pair-wise choices of constants given a $\boldsymbol{k}$ vector. Equation (2) was originally proven in (Beame et al. 2015), we refer the reader to (Malhotra and Serafini 2021) for the formulation presented here.

Families of Probability Distributions and Projectivity. We will be interested in probability distributions over the set of interpretations or equivalently their multi-relational graphs. A family of probability distributions $\{P^{(n)} : n \in \mathbb{N}\}$ specifies, for each finite domain of size n, a distribution $P^{(n)}$ on the possible n-world set $\Omega^{(n)}$ (Jaeger and Schulte 2020). We will mostly work with the so-called exchangeable probability distributions (Jaeger and Schulte 2020) i.e. distributions where $P^{(n)}(\omega) = P^{(n)}(\omega')$ if ω and ω' are isomorphic. A distribution $P^{(n)}(\omega)$ over n-worlds induces a marginal probability distribution over m-worlds $\omega' \in \Omega^{(m)}$ as follows:

$$P^{(n)} \downarrow [m](\omega') = \sum_{\omega \in \Omega^{(n)} : \omega \downarrow [m] = \omega'} P^{(n)}(\omega)$$

Notice that due to exchangeability $P^{(n)} \downarrow \mathrm{I}$ is the same for all subsets I of size m, hence we can always assume any induced m-world to be $\omega \downarrow [m]$. We are now able to define projectivity as follows:

Definition 1 ((Jaeger and Schulte 2020))**.** *An exchangeable family of probability distributions is called projective if for all $m < n$:*

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

When dealing with probability distributions over multi-relational representation, we denote by $(\boldsymbol{X}, \boldsymbol{Y})$ the random vector where, $\boldsymbol{X} = (X_1, \ldots, X_n)$ and each X_i takes value in $[u]$; and $\boldsymbol{Y} = (Y_{12}, Y_{13}, \ldots, Y_{qr}, \ldots, Y_{n-1,n})$ where $q < r$ and Y_{qr} takes values in $[b]$.

A Parametric Normal Form for MLNs

A Markov Logic Network (MLN) Φ is defined by a set of weighted formulas $\{(\phi_i, a_i)\}_i$, where ϕ_i are quantifier free, function-free FOL formulas with weights $a_i \in \mathbb{R}$. An MLN Φ induces a probability distribution over the set of possible worlds $\omega \in \Omega^{(n)}$:

$$P_{\Phi}^{(n)}(\omega) = \frac{1}{Z(n)} \exp\Big(\sum_{(\phi_i, a_i) \in \Phi} a_i . N(\phi_i, \omega)\Big)$$

where $N(\phi_i, \omega)$ represents the number of true groundings of ϕ_i in ω. The normalization constant $Z(n)$ is called the *partition function* that ensures that $P^{(n)}_{\Phi}$ is a probability distribution.

Theorem 1. *Any Markov Logic Network (MLN) $\Phi = \{(\phi_i, a_i)\}_i$ on a domain of size n, such that ϕ_i contains at-most two variables, can be expressed as follows:*

$$P^{(n)}_{\Phi}(\omega) = \frac{1}{Z(n)} \prod_{i \in [u]} s_i^{k_i} \prod_{\substack{i,j \in [u] \\ i \leq j}} \prod_{l \in [b]} (t_{ijl})^{h^{ij}_l} \tag{4}$$

where s_i and t_{ijl} are positive real numbers and k_i is $k_i(\omega)$ and h^{ij}_l is equal to $h^{ij}_l(\omega)$.

Proof. Let $\Phi = \{(\phi_i, a_i)\}_i$ be an MLN, such that ϕ_i contains at-most two variables. Firstly, every weighted formula $(\phi(x,y), a) \in \Phi$ that contains exactly two variables is replaced by two weighted formulas $(\phi(x,x), a)$ and $(\phi(x,y) \wedge (x \neq y), a)$. The MLN distribution $P^{(n)}_{\Phi}$ is invariant under this transformation. Hence, Φ can be equivalently written as $\{(\alpha_q(x), a_q)\}_q \cup \{(\beta_p(x,y), b_p)\}_p$, where $\{\alpha_q(x)\}_q$ is the set of formulas containing only the variable x and $\{\beta_p(x,y)\}_p$ is the set of formulas containing both the variables x and y. Notice that every $\beta_p(x,y)$ entails $x \neq y$.

Let us have $\omega \in \Omega^{(n)}$, where we have a domain constant c such that $\omega \models i(c)$. Now notice that the truth value of ground formulas $\{\alpha_q(c)\}_q$ in ω is completely determined by $i(c)$ irrespective of all other domain constants. Hence, the (multiplicative) weight contribution of $i(c)$ to the weight of ω can be given as $\exp(\sum_q a_q \mathbb{1}_{i(x) \models \alpha_q(x)})$. We define s_i as follows:

$$s_i = \exp(\sum_q a_q \mathbb{1}_{i(x) \models \alpha_q(x)}) \tag{5}$$

Clearly, this argument can be repeated for all the domain constants realizing any 1-type in $[u]$. Hence, the (multiplicative) weight contribution due to 1-types of all domain constants and equivalently due to the groundings of all unary formulas, is given as $\prod_{i \in [u]} s_i^{k_i}$.

We are now left with weight contributions due to the binary formulas $\{(\beta_p(x,y), b_p)\}_p$. Due to the aforementioned transformation, each binary formula $\beta(x,y)$ contains a conjunct $(x \neq y)$. Hence, all groundings of $\beta(x,y)$ such that both x and y are mapped to the same domain constants evaluate to false. Hence, we can assume that x and y are always mapped to distinct domain constants. Let us have an unordered pair of domain constants $\{c, d\}$ such that $\omega \models ijl(c,d)$. The truth value of any binary ground formula $\beta(c,d)$ and $\beta(d,c)$ is completely determined by $ijl(c,d)$ irrespective of all other domain constants. Hence, the multiplicative weight contribution due to the ground formulas $\{\beta_p(c,d)\}_p \cup \{\beta_p(d,c)\}_p$ is given as t_{ijl}, where t_{ijl} is defined as follows:

$$\exp\left(\sum_p b_p \mathbb{1}_{ijl(x,y) \models \beta_p(x,y)} + \sum_p b_p \mathbb{1}_{ijl(x,y) \models \beta_p(y,x)}\right) \tag{6}$$

Hence, the weight of an interpretation ω under the MLN Φ is given as

$$\prod_{i \in [u]} s_i^{k_i} \prod_{\substack{i,j \in [u] \\ i \leq j}} \prod_{l \in [b]} (t_{ijl})^{h^{ij}_l}$$

□

Definition 2. *Given an MLN in the parametric normal form given by equation* (4)*. Then f_{ij} is defined as $\sum_{l \in [b]} t_{ijl}$.*

We will now provide the parameterized version of the partition function $Z(n)$ due to Theorem 1.

Proposition 1. *Let Φ be an MLN in the form* (4)*, then the partition function $Z(n)$ is given as:*

$$Z(n) = \sum_{\boldsymbol{k}} \binom{n}{\boldsymbol{k}} \prod_{i \in [u]} s_i^{k_i} \prod_{\substack{i,j \in [u] \\ i \leq j}} (f_{ij})^{\boldsymbol{k}(i,j)} \tag{7}$$

where $\boldsymbol{k}(i,j)$ is defined in equation (3)*.*

Sketch. The proposition is a parameterized version of equation (2), where $\prod_{i \in [u]} s_i^{k_i}$ takes into account the weight contributions due to the 1-type realizations and f_{ij} is essentially a weighted version of n_{ij} i.e. given a pair of constants c and d such that they realize the i^{th} and the j^{th} 1-type respectively, then f_{ij} is the sum of the weights due to the 2-types realized by the extensions to the binary predicates containing both c and d.

□

Projectivity in Markov Logic Networks

We present the necessary and sufficient conditions for an MLN to be projective in the two variable fragment. The complete proofs are provided in the appendix.

Lemma 1 (Sufficiency)**.** *A Markov Logic Network in the two variable fragment is projective if all the f_{ij} have the same value i.e. $\forall i, j \in [u] : f_{ij} = F$, for some positive real number F.*

Sketch. The key idea of the proof is that if $\forall i, j \in [u] : f_{ij} = F$, then the partition function factorizes as $Z(n) = (F)^{\binom{n}{2}} \left(\sum_{i \in [u]} s_i\right)^n$. Now, defining $p_i = \frac{s_i}{\sum_i s_i}$ and $w_{ijl} = \frac{t_{ijl}}{F}$, allows us to re-define the MLN distribution (4) equivalently as follows:

$$P^{(n)}_{\Phi}(\omega) = \prod_{i \in [u]} p_i^{k_i} \prod_{\substack{i,j \in [u] \\ i \leq j}} \prod_{l \in [b]} w_{ijl}^{h^{ij}_l} \tag{8}$$

Here, $\sum_i p_i = 1$ and $\sum_l w_{ijl} = 1$. Hence, $P^{(n)}_{\Phi}(\omega)$ is essentially a (labeled) stochastic block model, which are known to be projective (Shalizi and Rinaldo 2013). □

We will now prove that the aforementioned sufficient conditions are also necessary.

Lemma 2 (Necessary)**.** *If a Markov Logic network in the two variable fragment is projective then, all the f_{ij} have the same value i.e. $\forall i, j \in [u] : f_{ij} = F$, for some positive real number F.*

Sketch. We begin by writing the projectivity condition in the multi-relational representation, i.e. $P_{\Phi}^{(n+1)} \downarrow [n](\boldsymbol{X}' = \boldsymbol{x}', \boldsymbol{Y}' = \boldsymbol{y}')$ is equal to:

$$\sum_{\substack{\boldsymbol{x}_{[n]}=\boldsymbol{x}' \\ \boldsymbol{y}_{[n]}=\boldsymbol{y}'}} P_{\Phi}^{(n+1)}(\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Y} = \boldsymbol{y}) \tag{9}$$

Multiplying and dividing equation (9) by $Z(n)$ and using simple algebraic manipulations we get that for all $\boldsymbol{x}'$:

$$\frac{Z(n+1)}{Z(n)} = \sum_{i \in [u]} s_i \prod_{j \in [u]} f_{ij}^{k_j(\boldsymbol{x}')} \tag{10}$$

Now, the LHS of equation (10) is completely independent of $\boldsymbol{x}'$, whereas RHS is dependent on $\boldsymbol{x}'$. It can be shown that this is possible iff f_{ij} does not depend on $\boldsymbol{x}'$, which in turn is possible iff f_{ij} does not depend on i and j i.e. $\forall i, j \in [u] : f_{ij} = F$, for some positive real F. □

We are finally able to provide the following theorem.

Theorem 2. *A Markov Logic Network (MLN) $\Phi = \{(\phi_i, a_i)\}_i$, such that ϕ_i contains at-most two variables is projective if and only if all the f_{ij} (as given in Definition 2) have the same value i.e. $\forall i, j \in [u] : f_{ij} = F$, for some positive real number F.*

In the next section, we will show that the conditions in Theorem 2 correspond to a special type of probability distributions. We will characterize such distributions and then investigate their properties.

Relational Block Model

In this section we introduce the Relational Block Model (RBM). We show that any projective MLN in the two variable fragment can be expressed as an RBM. And any RBM can be expressed as a projective MLN. Furthermore, we show that an RBM is a unique characterization of a projective MLN in the two variable fragment.

Definition 3. *Let n be a positive integer (the number of domain constants), u be a positive integer (the number of 1-types), b be a positive integer (the number of 2-tables), $p = (p_1, ..., p_u)$ be a probability vector on $[u] = \{1, ..., u\}$ and $W = (w_{ijl}) \in [0,1]^{u \times u \times b}$, where $w_{ijl} = w_{\overline{ijl}}$ (w_{ijl} is the conditional probability of domain elements (c, d) realizing the l^{th} 2-table, given $i(c)$ and $j(d)$). The multi-relational graph $(\boldsymbol{x}, \boldsymbol{y})$ is drawn under $\mathrm{RBM}(n, p, W)$ if $\boldsymbol{x}$ is an n-dimensional vector with $i.i.d$ components distributed under p and $\boldsymbol{y}$ is a random vector with its component $y_{qr} = l$, where $l \in [b]$, with a probability $w_{x_q x_r l}$ independently of all other pair of domain constants.*

Thus, the probability distribution of $(\boldsymbol{x}, \boldsymbol{y})$ is defined as follows, where $\boldsymbol{x} \in [u]^n$ and $\boldsymbol{y} \in [b]^{\binom{n}{2}}$

$$\begin{aligned}
P(\boldsymbol{X} = \boldsymbol{x}) &:= \prod_{q=1}^{n} p_{x_q} = \prod_{i=1}^{u} p_{x_i}^{k_i} \\
P(\boldsymbol{Y} = \boldsymbol{y} | \boldsymbol{X} = \boldsymbol{x}) &:= \prod_{1 \leq q < r \leq n} w_{x_q x_r y_{qr}} \\
&= \prod_{1 \leq i \leq j \leq u} \prod_{1 \leq l \leq b} (w_{ijl})^{h_l^{ij}}
\end{aligned}$$

In the following example, we show how RBMs can model homophily.

Example 2 (Homophily)**.** *Let us have an FO^2 language with a unary predicate C (representing a two colors) and a binary predicate R. We wish to model a distribution on simple undirected graphs i.e. models of the formula $\phi = \forall xy.\neg R(x, x) \land (R(x, y) \rightarrow R(y, x))$ such that same color nodes are more likely to have an edge. Due to ϕ the 1-types with $\neg R(x, x)$ as a conjunct have a probability zero. Hence, we can assume we have only two 1-types: $1(x) = C(x) \land \neg R(x, x)$ and $2(x) = \neg C(x) \land \neg R(x, x)$ (representing two possible colors for a given node). Similarly due to ϕ, we have only two 2-tables $1(x, y) : R(x, y) \land R(y, x)$ and $2(x, y) : \neg R(x, y) \land \neg R(y, x)$ (representing existence and non existence of edges). We can now easily define homophily by following parameterization of an RBM. $p_1 = p_2 = 0.5$ i.e. any node can have two colors with equal probability. Then we can define $w_{111} = 0.9$, $w_{112} = 0.1$, $w_{221} = 0.9$, $w_{222} = 0.1$, $w_{121} = 0.1$ and $w_{122} = 0.9$.*

Theorem 3. *Every projective Markov Logic Network in the two variable fragment can be expressed as an RBM.*

Proof. The proof follows from the sufficiency proof in Lemma 3. Notice that in the proof, we derive equation (8) (equivalently, equation (21) in the appendix), which is exactly the expression for RBM. Hence, any projective MLN can be converted to an RBM by defining p_i and w_{ijl} as follows:

$$p_i = \frac{s_i}{\sum_i s_i} \quad w_{ijl} = \frac{t_{ijl}}{\sum_l t_{ijl}} \tag{11}$$

□

Theorem 4. *Every RBM can be expressed as a projective MLN in the two variable fragment.*

Proof. Given an RBM as defined in definition 3 with parameters $\{p_i, w_{ijl}\}$, let us have a projective MLN Φ such that every 1-type $i(x)$ is a formula in the MLN with a weight $\log p_i$. Φ also has a weighted formula $ijl(x, y)$ for every 2-type, such that $i \leq j$. The weight for $ijl(x, y)$ is $\log(w_{ijl})$ if $ijl(x, y) \neq ijl(y, x)$, and is $0.5 \log(w_{ijl})$ if $ijl(x, y) = ijl(y, x)$. It can be seen from definition of s_i (5) and t_{ijl} (6), that for Φ, $s_i = p_i$ and $t_{ijl} = w_{ijl}$. Hence, due to (4), we have that:

$$P_{\Phi}^{(n)}(\omega) = \frac{1}{Z(n)} \prod_{i \in [u]} p_i^{k_i} \prod_{\substack{i,j \in [u] \\ i \leq j}} \prod_{l \in [b]} (w_{ijl})^{h_l^{ij}} \tag{12}$$

In the MLN Φ, $\sum_i s_i = \sum_i p_i = 1$ and $\sum_l t_{ijl} = \sum_l w_{ijl} = f_{ij} = 1$. Hence, using Proposition 1, we have that $Z(n) = 1$. Hence, completing the proof.

$\square$

Proposition 2. *Given two RBMs with probability distribution P' and P'' and parameters $\{p'_i, w'_{ijl}\}$ and $\{p''_i, w''_{ijl}\}$. If $P' = P''$, then, $p'_i = p''_i$ and $w'_{ijl} = w''_{ijl}$.*

Proof. The proposition is a consequence of the fact that the parameter p_i is marginal probability of an arbitrary constant c realizing the i^{th} 1-type and w_{ijl} is the conditional probability of an arbitrary pair of constants (c, d) realizing the l^{th} 2-table given $i(c)$ and $j(d)$. Hence, two RBMs that disagree on the p_i and w_{ijl} cannot assign the same probability mass to marginal probability of $i(c)$ and $ijl(c, d)$ and hence, cannot be the same distribution.

$\square$

Corollary 1 (of Proposition 2)**.** *Given two projective MLNs Φ' and Φ'' such that they have the same probability distributions $P_{\Phi'}$ and $P_{\Phi''}$, with there respective RBMs parameterized by $\{p'_i, w'_{ijl}\}$ and $\{p''_i, w''_{ijl}\}$. Then we must have that $p'_i = p''_i$ and $w'_{ijl} = w''_{ijl}$.*

Hence, RBMs are a unique representation for projective MLNs in the two variable fragment.

Previous Characterizations of Projectivity

Jaeger and Schulte (Jaeger and Schulte 2018) show that an MLN is projective if it's formulae ϕ_i satisfy the property that any two atoms appearing in ϕ_i contain exactly the same variables. Such MLNs are also known as *σ-determinate* (Singla and Domingos 2012). We now show that in the two variable fragment, Theorem 2 leads to a strictly more expressive class of MLNs.

Proposition 3. *Given an MLN $\Phi = \{\phi_i, a_i\}_i$ such that any two atoms appearing in ϕ_i contain exactly the same variables or equivalently that the MLN is $\sigma-$determinate. Then:*

$$\forall i, j, i', j' \in [u], \forall l \in [b] : t_{i'j'l} = t_{ijl} \tag{13}$$

Proof. We first write an equivalent MLN $\Phi' = \{\alpha_q(x), a_q\} \cup \{\beta_p(x, y), b_p\}$ as presented in proof of Theorem 1. Due to the conditions provided in the proposition, all the atoms in $\beta_p(x, y)$ contain both the variables x and y. Using the definition of t_{ijl} from (6), and the fact that none of the $\beta_p(x, y)$ have an atom with only one variable, we have that the value of t_{ijl} depends only on the l^{th} 2-table, irrespective of the 1-types i and j. This is because, none of the first order atoms in the i^{th} and the j^{th} 1-type appear in $\beta_p(x, y)$. Hence, t_{ijl} only depends on l, giving us equation (13). $\square$

Proposition 3 is a stricter condition than Theorem 2. In the following, we prove that σ-determinate MLNs cannot express all the projective MLNs in the two variable fragment.

Theorem 5. *There exists a projective MLN in the two variable fragment which cannot be expressed as a $\sigma-$determinate MLN.*

Proof. Let us have a σ-determinate MLN Φ, since Φ is projective, we can create it's equivalent RBM (due to Theorem 4), say P. Let $\{p_i, w_{ijl}\}$ be the parameters of P. Due to equation (11) and Proposition 3, we have that $w_{ijl} = w_{i'j'l}$ for all i, j, i', j'. Due to existence of a projective MLN for every RBM (from Theorem 4), we can always create an MLN Φ' for which the RBM parameters $w_{ijl} \neq w_{i'j'l}$ for some i, j, i', j'. Since, RBMs uniquely characterize the probability distributions due to MLNs (from Corollary 1), Φ' can not be expressed as an MLN such that $w_{ijl} = w_{i'j'l}$. Hence, Φ' can not be expressed as a σ-determinate MLN.

$\square$

In the following example, we provide an MLN which cannot be written as a $\sigma-$determinate MLN.

Example 3. *Let us have a binary predicate R. We have only two 1-types $R(x, x)$ (say $1(x)$) and $\neg R(x, x)$ (say $2(x)$) and four 2-tables, $R(x, y) \wedge R(y, x)$ (say $1(x, y)$), $R(x, y) \wedge \neg R(y, x)$ (say $2(x, y)$), $\neg R(x, y) \wedge R(y, x)$ (say $3(x, y)$) and $\neg R(x, y) \wedge \neg R(y, x)$ (say $4(x, y)$). An MLN Φ, with the following 2-types as weighted formulas, cannot be expressed as a $\sigma-$determinate MLN:*

$$111(x, y) : \log 7 \quad 114(x, y) : \log 4$$
$$124(x, y) : \log 64 \quad 221(x, y) : \log 8$$

In parametric normal form, $t_{111} = \exp(2 \log 7)$, $t_{114} = \exp(2 \log 4)$, $t_{124} = \exp(\log 64)$ and $t_{221} = \exp(2 \log 8)$. All the other t_{ijl}, such that $ijl(x, y)$ is not a dual of $111(x, y), 114(x, y), 124(x, y)$ or $221(x, y)$, are equal to $\exp(0)$ i.e. 1. It can be verified that $f_{ij} = 67$ for all $i, j \in [2]$, hence, this MLN is projective due to Theorem 2. Using Theorem 3, we can express this distribution as an RBM, such that $w_{111} = \frac{7^2}{67}$ and $w_{114} = \frac{4^2}{67}$. If $w_{111} \neq w_{114}$ then necessarily $t_{111} \neq t_{114}$ (as w_{ijl} is defined as $\frac{t_{ijl}}{f_{ij}}$ and f_{ij} is the same for all i, j in Φ and in any equivalent MLN, due to Theorem 2). Due to uniqueness of RBM parameters for any set of projective MLNs expressing the same distribution (Corollary 1), we have that in all MLNs equivalent to Φ, $t_{111} \neq t_{114}$. Hence, using Proposition 3, we have that any MLN expressing the same distribution as Φ cannot be expressed as a $\sigma-$determinate MLN.

Maximum Likelihood Learning

In a learning setting, for an MLN $\{\phi_i, a_i\}$ in the two variable fragment, we are interested in estimating the set of parameters $\boldsymbol{\theta} = \{a_i\}$ that maximize the likelihood of a training example such that the learnt MLN is projective. As analyzed in (Xiang and Neville 2011; Kuzelka et al. 2017), we will focus on the scenario where only a single possible world $\omega \in \Omega^{(n)}$ is observed. We estimate $\boldsymbol{\theta}$ by maximizing the likelihood

$$L^{(n)}(\boldsymbol{\theta}|\omega) = P^{(n)}_{\boldsymbol{\theta}}(\omega) \tag{14}$$

Notice that although every projective MLN can be equivalently defined as an RBM, the maximum likelihood parameter estimate for an RBM is not the same as the parameter estimate for an MLN such that it is projective.

We will now provide, the maximum likelihood estimator for an RBM.

Proposition 4. *Given a training example $\omega \in \Omega^{(n)}$, the maximum likelihood parameter estimate for an RBM is given as, $p_i = \frac{k_i}{n}$ and $w_{ijl} = \frac{h_l^{ij}}{\boldsymbol{k}(i,j)}$.*

Proposition 4 can be derived by maximizing the log likelihood due to the distribution given in Definition 3.

We will now see how maximum likelihood parameter estimate can be obtained for an MLN such that the MLN is projective.

Given an MLN $\{\phi_i, a_i\}_i$ in the two variable fragment, where $\boldsymbol{\theta} = \{a_i\}_i$ are unknown parameters to be estimated, due to Theorem 1, we can define $s_i(\boldsymbol{\theta})$ and $t_{ijl}(\boldsymbol{\theta})$, such that the likelihood is given as:

$$L(\boldsymbol{\theta}|\omega) = \frac{1}{Z(n)} \prod_{i \in [u]} s_i(\boldsymbol{\theta})^{k_i} \prod_{\substack{i,j \in [u] \\ i \leq j}} \prod_{l \in [b]} (t_{ijl}(\boldsymbol{\theta}))^{h_l^{ij}} \tag{15}$$

Defining $F(\boldsymbol{\theta})$ as $\sum_l t_{i'j'l}(\boldsymbol{\theta})$ for some fixed i' and j', the maximum likelihood parameter estimates such that the estimated MLN is projective, can be then obtained by solving the following optimization problem:

$$\begin{aligned} \underset{\boldsymbol{\theta}}{maximize} : \Big[&\sum_{i \in [u]} k_i \log s_i(\boldsymbol{\theta}) + \sum_{\substack{i,j \in [u] \\ i \leq j}} \sum_{l \in [b]} h_l^{ij} \log t_{ijl}(\boldsymbol{\theta}) \\ &- n \log \big(\sum_{i \in [u]} s_i(\boldsymbol{\theta})\big) - \binom{n}{2} \log F(\boldsymbol{\theta}) \Big] \\ subject \quad to : \quad &\forall i, j \in [u] : f_{ij}(\boldsymbol{\theta}) = F(\boldsymbol{\theta}) \end{aligned} \tag{16}$$

Notice that due to factorization of $Z(n)$ under projectivity (see Lemma 3), $-n \log \big(\sum_{i \in [u]} s_i(\boldsymbol{\theta})\big) - \binom{n}{2} \log F(\boldsymbol{\theta})$ represents $-\log(Z(n))$. The above optimization can be solved through any conventional optimization algorithm. It can be seen that this problem has a much lesser overhead as far as computing $\log(Z(n))$ is concerned. But the additional constraints may counter act this gain. Furthermore, in many cases it may happen that no non-zero weights exist that satisfy the constraints and in that case the problem will return zero weights for the MLN formulas.

Theorem 6. *Given a training example $\omega \in \Omega^{(n)}$, then there is no parameterization for any projective MLN in the two variable fragment that has a higher likelihood for ω than the maximum likelihood RBM for ω.*

Proof. Let L be the likelihood of ω due to the maximum likelihood RBM. Let L' be the likelihood of ω due to a projective MLN Φ, such that $L' > L$. Now, due to Theorem 3, Φ can be expressed as an RBM. Hence, we can have an RBM such that the likelihood of ω is L', but $L' > L$ which is a contradiction. Hence, we cannot have a projective MLN that gives a higher likelihood to ω than the maximum likelihood RBM. □

Theorem 6 shows us that if a data source is known to be projective (i.e. we know that marginals in the data will be independent of the domain at large) then in terms of likelihood, specially in the case of large relational datasets, we are better off in using an RBM than an expert defined MLN. This can also be argued from efficiency point of view as RBMs admit much more efficient parameter estimates.

We will now move on to the question: *are parameters learned on a domain of size n, also good for modelling domain of a different size m ?* This question is an abstraction of many real world problems, for example, learning over relational data in presence of incomplete information (Kossinets 2006), modelling a social network from only sub-sampled populations (Handcock and Gile 2010), modelling progression of a disease in a population by only testing a small set of individuals (Srinivasavaradhan et al. 2021) etc.

Jaeger and Schulte (Jaeger and Schulte 2018) formalized the afore mentioned notions in the following two criterions:

$$E_\omega[\underset{\boldsymbol{\theta}}{\text{argmax}} \log L^{(m)}(\boldsymbol{\theta}|\omega')] = \underset{\boldsymbol{\theta}}{\text{argmax}} \log L^{(n)}(\boldsymbol{\theta}|\omega) \tag{17}$$

$$\underset{\boldsymbol{\theta}}{\text{argmax}} E_\omega[\log L^{(m)}(\boldsymbol{\theta}|\omega')] = \underset{\boldsymbol{\theta}}{\text{argmax}} \log L^{(n)}(\boldsymbol{\theta}|\omega) \tag{18}$$

It is easy to see, by law of large numbers, that RBMs satisfy both these criterions. On the other hand the same can not be said about the maximum likelihood estimates for projective MLNs as described in (16).

Conclusion

In this work, we have characterized the class of projective MLNs in the two-variable fragment. We have also identified a special class of models, namely Relational Block Model. We show that the maximum likelihood RBM maximizes the training data likelihood w.r.t to any projective MLN in the two-variable fragment. Furthermore, RBMs admit consistent parameter learning from sub-sampled domains, potentially allowing them to scale to very large datasets, especially in situations where the test data size is not known or changes over time.

From an applications point of view, the superiority of RBMs in terms of training likelihood maximization and consistent parameter learning can potentially make them a better choice over an expert defined MLN, especially when training set is large and the test domain size is unknown or varies over time. We plan to investigate such capabilities of RBMs and projective MLNs in future work, especially in comparison to models like Adaptive MLNs (Jain, Barthels, and Beetz 2010) and Domain Size Aware MLNs (Mittal et al. 2019).

On the theoretical front, the imposed independence structure due to projectivity clearly resembles the AHK models proposed in (Jaeger and Schulte 2020). In future works, we aim at investigating this resemblance and generalizing our work to capture complete projectivity criterion for all the MLNs.

Acknowledgements

We would like to thank Manfred Jaeger and Felix Weitkämper for their valuable critique and discussion time on the topic.

References

Beame, P.; den Broeck, G. V.; Gribkoff, E.; and Suciu, D. 2015. Symmetric Weighted First-Order Model Counting. In Milo, T.; and Calvanese, D., eds., *Proceedings of the 34th ACM Symposium on Principles of Database Systems, PODS 2015, Melbourne, Victoria, Australia, May 31 - June 4, 2015*, 313–328. ACM.

Getoor, L.; and Taskar, B. 2007. *Introduction to Statistical Relational Learning (Adaptive Computation and Machine Learning)*. The MIT Press. ISBN 0262072882.

Handcock, M. S.; and Gile, K. J. 2010. Modeling social networks from sampled data. *The Annals of Applied Statistics*, 4(1): 5 – 25.

Jaeger, M.; and Schulte, O. 2018. Inference, Learning, and Population Size: Projectivity for SRL Models. *CoRR*, abs/1807.00564.

Jaeger, M.; and Schulte, O. 2020. A Complete Characterization of Projectivity for Statistical Relational Models. In Bessiere, C., ed., *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, 4283–4290. ijcai.org.

Jain, D.; Barthels, A.; and Beetz, M. 2010. Adaptive Markov Logic Networks: Learning Statistical Relational Models with Dynamic Parameters. In *Proceedings of the 2010 Conference on ECAI 2010: 19th European Conference on Artificial Intelligence*, 937–942. NLD: IOS Press. ISBN 9781607506058.

Kossinets, G. 2006. Effects of missing data in social networks. *Social Networks*, 28(3): 247–268.

Kuusisto, A.; and Lutz, C. 2018. Weighted model counting beyond two-variable logic. In Dawar, A.; and Grädel, E., eds., *Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2018, Oxford, UK, July 09-12, 2018*, 619–628. ACM.

Kuzelka, O.; Wang, Y.; Davis, J.; and Schockaert, S. 2017. Relational Marginal Problems: Theory and Estimation. *CoRR*, abs/1709.05825.

Malhotra, S.; and Serafini, L. 2021. Weighted Model Counting in FO2 with Cardinality Constraints and Counting Quantifiers: A Closed Form Formula. To Appear in Proceedings of AAAI 2022.

Mittal, H.; Bhardwaj, A.; Gogate, V.; and Singla, P. 2019. Domain-Size Aware Markov Logic Networks. In Chaudhuri, K.; and Sugiyama, M., eds., *The 22nd International Conference on Artificial Intelligence and Statistics, AISTATS 2019, 16-18 April 2019, Naha, Okinawa, Japan*, volume 89 of *Proceedings of Machine Learning Research*, 3216–3224. PMLR.

Poole, D.; Buchman, D.; Kazemi, S. M.; Kersting, K.; and Natarajan, S. 2014. Population Size Extrapolation in Relational Probabilistic Modelling. In Straccia, U.; and Calì, A., eds., *Scalable Uncertainty Management - 8th International Conference, SUM 2014, Oxford, UK, September 15-17, 2014. Proceedings*, volume 8720 of *Lecture Notes in Computer Science*, 292–305. Springer.

Raedt, L. D.; Kersting, K.; Natarajan, S.; and Poole, D. 2016. *Statistical Relational Artificial Intelligence: Logic, Probability, and Computation*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers.

Richardson, M.; and Domingos, P. 2006. Markov logic networks. *Machine learning*, 62(1-2): 107–136.

Shalizi, C. R.; and Rinaldo, A. 2013. CONSISTENCY UNDER SAMPLING OF EXPONENTIAL RANDOM GRAPH MODELS. *Annals of statistics*, 41 2: 508–535.

Singla, P.; and Domingos, P. M. 2012. Markov Logic in Infinite Domains. *CoRR*, abs/1206.5292.

Snijders, T. A. B. 2010. Conditional Marginalization for Exponential Random Graph Models. *The Journal of Mathematical Sociology*, 34(4): 239–252.

Srinivasavaradhan, S. R.; Nikolopoulos, P.; Fragouli, C.; and Diggavi, S. 2021. Dynamic group testing to control and monitor disease progression in a population. arXiv:2106.10765.

Weitkämper, F. Q. 2021. An Asymptotic Analysis of Probabilistic Logic Programming, with Implications for Expressing Projective Families of Distributions. *Theory Pract. Log. Program.*, 21(6): 802–817.

Xiang, R.; and Neville, J. 2011. Relational Learning with One Network: An Asymptotic Analysis. In Gordon, G. J.; Dunson, D. B.; and Dudík, M., eds., *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2011, Fort Lauderdale, USA, April 11-13, 2011*, volume 15 of *JMLR Proceedings*, 779–788. JMLR.org.

Appendix

A.1 : Lemma 3 [Sufficiency]

Lemma 3 (Sufficiency)**.** *A Markov Logic Network in the two variable fragment is projective if all the f_{ij} have the same value i.e. $\forall i, j \in [u] : f_{ij} = F$, for some positive real number F.*

Proof. Let $\forall i, j \in [u] : f_{ij} = F$. Hence, due to Proposition 1, we have:

$$Z(n) = \sum_{\boldsymbol{k}} \binom{n}{\boldsymbol{k}} \prod_{i \in [u]} s_i^{k_i} \prod_{\substack{i,j \in [u] \\ i \leq j}} (F)^{\boldsymbol{k}(i,j)} \tag{19}$$

$$= \sum_{\boldsymbol{k}} \binom{n}{\boldsymbol{k}} \prod_{i \in [u]} s_i^{k_i} (F)^{\binom{n}{2}} = F^{\binom{n}{2}} \Big(\sum_{i \in [u]} s_i \Big)^n \tag{20}$$

Let $p_i = \left(\frac{s_i}{\sum_i s_i} \right)$ and $w_{ijl} = \left(\frac{t_{ijl}}{F} \right)$. Hence,

$$P_{\Phi}^{(n)}(\omega) = \frac{1}{(\sum_{i \in [u]} s_i)^n (F)^{\binom{n}{2}}} \prod_{i \in [u]} s_i^{k_i} \prod_{\substack{i,j \in [u] \\ i \leq j}} \prod_{l \in [b]} (t_{ijl})^{h_l^{ij}}$$

$$= \prod_{i \in [u]} \Big(\frac{s_i}{\sum_{i \in [u]} s_i} \Big)^{k_i} \prod_{\substack{i,j \in [u] \\ i \leq j}} \prod_{l \in [b]} (\frac{t_{ijl}}{F})^{h_l^{ij}}$$

$$= \prod_{i \in [u]} p_i^{k_i} \prod_{\substack{i,j \in [u] \\ i \leq j}} \prod_{l \in [b]} w_{ijl}^{h_l^{ij}}$$

Using the multi-relational representation, $P_{\Phi}^{(n)}(\omega)$ can be equivalently expressed as:

$$P_{\Phi}^{(n)}(\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Y} = \boldsymbol{y}) = \prod_{q \in [n]} p_{x_q} \prod_{\substack{q,r \in [n] \\ q < r}} w_{x_q x_r y_{qr}} \tag{21}$$

Let $(\boldsymbol{X}', \boldsymbol{Y}')$ be the random vector containing X_q and $Y_{p,q}$ with $p < q \in [m]$. Clearly, our goal is to show that

$$P_{\Phi}^{(n)} \downarrow [m](\boldsymbol{X}' = \boldsymbol{x}', \boldsymbol{Y}' = \boldsymbol{y}') = P_{\Phi}^{(m)}(\boldsymbol{X}' = \boldsymbol{x}', \boldsymbol{Y}' = \boldsymbol{y}')$$

Now, the marginal distribution over the m-worlds $(\boldsymbol{X}', \boldsymbol{Y}')$, due to $P_{\Phi}^{(n)}(\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Y} = \boldsymbol{y})$ can be expressed as:

$$P_{\Phi}^{(n)} \downarrow [m](\boldsymbol{X}' = \boldsymbol{x}', \boldsymbol{Y}' = \boldsymbol{y}') = \sum_{\substack{\boldsymbol{x}_{[m]} = \boldsymbol{x}' \\ \boldsymbol{y}_{[m]} = \boldsymbol{y}'}} P_{\Phi}^{(n)}(\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Y} = \boldsymbol{y})$$

$$= \sum_{\substack{\boldsymbol{x}_{[m]} = \boldsymbol{x}' \\ \boldsymbol{y}_{[m]} = \boldsymbol{y}'}} \prod_{q \in [n]} p_{x_q} \prod_{\substack{q,r \in [n] \\ q < r}} w_{x_q x_r y_{qr}}$$

$$= \prod_{q \in [m]} p_{x_q} \prod_{\substack{q,r \in [m] \\ q < r}} w_{x_q x_r y_{qr}} \times \left(\sum_{\substack{\boldsymbol{x}_{[\overline{m}]} \\ \boldsymbol{y}_{[\overline{m}]}}} \prod_{q \in [\overline{m}]} p_{x_q} \prod_{\substack{q,r \in [\overline{m}] \\ q < r}} w_{x_q x_r y_{qr}} \prod_{\substack{q \in [m] \\ r \in [\overline{m}]}} w_{x_q x_r y_{qr}} \right)$$

$$= \prod_{i \in [u]} p_i^{k_i(\boldsymbol{x}')} \prod_{\substack{i,j \in [u] \\ i \leq j}} \prod_{l \in [b]} w_{ijl}^{h_l^{ij}(\boldsymbol{y}')}$$

$$\times \left(\sum_{\substack{\boldsymbol{x}_{[\overline{m}]} \\ \boldsymbol{y}_{[\overline{m}]}}} \prod_{i \in [u]} p_i^{k_i(\boldsymbol{x}_{[\overline{m}]})} \prod_{\substack{i,j \in [u] \\ i \leq j}} \prod_{l \in [b]} w_{ijl}^{h_l^{ij}(\boldsymbol{y}_{[\overline{m}]})} \prod_{\substack{i,j \in [u] \\ i \leq j}} \prod_{l \in [b]} w_{ijl}^{h_l^{ij}([m] \otimes [\overline{m}])} \right)$$

where $A \otimes B = A \times B \cup B \times A$. Notice that $\prod_{i\in[u]} p_i^{k_i(\boldsymbol{x'})} \prod_{\substack{i,j\in[u]\\ i\leq j}} \prod_{l\in[b]} w_{ijl}^{h_l^{ij}(\boldsymbol{y'})}$ is $P_{\Phi}^{(m)}(\boldsymbol{X'} = \boldsymbol{x'}, \boldsymbol{Y'} = \boldsymbol{y'})$. Hence, in order to complete the proof, we will now show that for any $\boldsymbol{x'}$:

$$\sum_{\substack{\boldsymbol{x}_{[\overline{m}]}\\ \boldsymbol{y}_{[\overline{m}]}}} \prod_{i\in[u]} p_i^{k_i(\boldsymbol{x}_{[\overline{m}]})} \prod_{\substack{i,j\in[u]\\ i\leq j}} \prod_{l\in[b]} w_{ijl}^{h_l^{ij}(\boldsymbol{y}_{[\overline{m}]})} \prod_{\substack{i,j\in[u]\\ i\leq j}} \prod_{l\in[b]} w_{ijl}^{h_l^{ij}([m]\otimes[\overline{m}])} = 1 \tag{22}$$

The LHS of equation (22) can be written as:

$$\sum_{\sum \boldsymbol{k}=n-m} \binom{n-m}{\boldsymbol{k}} \prod_{i\in[u]} p_i^{k_i} \prod_{\substack{i,j\in[u]\\ i\leq j}} (\sum_l w_{ijl})^{\boldsymbol{k}(i,j)} \prod_{\substack{i,j\in[u]\\ i\leq j}} (\sum_l w_{ijl})^{k_i(\boldsymbol{x'})\times k_j} \tag{23}$$

By definition, for any $i, j \in [u]$, $\sum_l w_{ijl} = 1$, and $\sum_i p_i = 1$. Hence, expression (23) can be written as:

$$\sum_{\sum \boldsymbol{k}=n-m} \binom{n-m}{\boldsymbol{k}} \prod_{i\in[u]} p_i^{k_i} = \left(\sum_i p_i\right)^{n-m} = 1$$

Hence, completing the proof. $\square$

A.2 : Lemma 4 [Necessary]

Lemma 4 (Necessary)**.** *If a Markov Logic network in the two variable fragment is projective then, all the f_{ij} have the same value i.e. $\forall i, j \in [u] : f_{ij} = F$, for some positive real number F.*

Proof. Let us have a markov logic network Φ over a domain $[n+1]$. Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be random vectors representing multi-relational graphs on the domain $[n+1]$. Let $\boldsymbol{X'}$ and $\boldsymbol{Y'}$ be random vectors representing multi-relational graphs on the domain $[n]$. Then :

$$P_{\Phi}^{(n+1)} \downarrow [n](\boldsymbol{X'} = \boldsymbol{x'}, \boldsymbol{Y'} = \boldsymbol{y'}) = \sum_{\substack{\boldsymbol{x}_{[n]}=\boldsymbol{x'}\\ \boldsymbol{y}_{[n]}=\boldsymbol{y'}}} P_{\Phi}^{(n+1)}(\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Y} = \boldsymbol{y})$$

$$= \sum_{\substack{\boldsymbol{x}_{[n]}=\boldsymbol{x'}\\ \boldsymbol{y}_{[n]}=\boldsymbol{y'}}} \frac{1}{Z(n+1)} \prod_{q\in[n+1]} s_{x_q} \prod_{\substack{q,r\in[n+1]\\ q<r}} t_{x_q x_r y_{qr}}$$

$$= \frac{1}{Z(n)} \prod_{q\in[n]} s_{x_q} \prod_{\substack{q,r\in[n]\\ q<r}} t_{x_q x_r y_{qr}} \frac{Z(n)}{Z(n+1)} \sum_{\substack{x_{n+1}\\ y_{q,n+1}}} s_{x_{n+1}} \prod_{q\in[n]} t_{x_q x_{n+1} y_{q,n+1}}$$

$$= P_{\Phi}^{(n)}(\boldsymbol{X'} = \boldsymbol{x'}, \boldsymbol{Y'} = \boldsymbol{y'}) \frac{Z(n)}{Z(n+1)} \sum_{\substack{x_{n+1}\\ y_{q,n+1}}} s_{x_{n+1}} \prod_{q\in[n]} t_{x_q x_{n+1} y_{q,n+1}}$$

Due to projectivity we have that:

$$P_{\Phi}^{(n)}(\boldsymbol{X'} = \boldsymbol{x'}, \boldsymbol{Y'} = \boldsymbol{y'}) = P_{\Phi}^{(n+1)} \downarrow [n](\boldsymbol{X'} = \boldsymbol{x'}, \boldsymbol{Y'} = \boldsymbol{y'})$$

Hence,

$$\frac{Z(n+1)}{Z(n)} = \sum_{\substack{x_{n+1}\\ y_{q,n+1}}} s_{x_{n+1}} \prod_{q\in[n]} t_{x_q x_{n+1} y_{q,n+1}}$$

which can be equivalently written as:

$$\frac{Z(n+1)}{Z(n)} = \sum_{i\in[u]} s_i \prod_{j\in[u]} \Big(\sum_{l\in[b]} t_{jil}\Big)^{k_j(\boldsymbol{x'})}$$

Now, $\sum_{l\in[b]} t_{jiv} = f_{ji} = f_{ij}$. Hence:

$$\frac{Z(n+1)}{Z(n)} = \sum_{i\in[u]} s_i \prod_{j\in[u]} f_{ij}^{k_j(\boldsymbol{x}')}$$

Hence, for any choice of the domain size m and for any choice of m-worlds $(\boldsymbol{x},\boldsymbol{y})$ and $(\boldsymbol{x}',\boldsymbol{y}')$, we have that:

$$\sum_{i\in[u]} s_i \prod_{j\in[u]} f_{ij}^{k_j(\boldsymbol{x})} = \sum_{i\in[u]} s_i \prod_{j\in[u]} f_{ij}^{k_j(\boldsymbol{x}')} \tag{24}$$

which implies [1] that:

$$\forall i, j, i', j' \in [u] : f_{ij} = f_{i'j'}$$

Hence, completing the proof. □

A.3 : Auxiliary Lemmas for Lemma 4

In proof of Lemma 4 we argue that, for any choice of the domain size m and for any choice of m-worlds $(\boldsymbol{x},\boldsymbol{y})$ and $(\boldsymbol{x}',\boldsymbol{y}')$, we have that:

$$\sum_{i\in[u]} s_i \prod_{j\in[u]} f_{ij}^{k_j(\boldsymbol{x})} = \sum_{i\in[u]} s_i \prod_{j\in[u]} f_{ij}^{k_j(\boldsymbol{x}')} \tag{25}$$

This implies that:

$$\forall i, j, i', j' \in [u] : f_{ij} = f_{i'j'} \tag{26}$$

We will first infer a slightly stricter equation from (25). $\boldsymbol{x}$ and $\boldsymbol{x}'$ can have any 1-type cardinalities, say $\boldsymbol{k} = \langle k_1(\boldsymbol{x})...k_u(\boldsymbol{x})\rangle = \langle k_1...k_u\rangle$ and $\boldsymbol{k}' = \langle k_1(\boldsymbol{x}')...k_u(\boldsymbol{x}')\rangle = \langle k_1'...k_u'\rangle$ respectively, such that $\sum_{i\in[u]} k_i = \sum_{i\in[u]} k_i' = m$. Hence, we can conclude that, for all $\boldsymbol{k}$ and $\boldsymbol{k}'$ such that $\sum_{i\in[u]} k_i = \sum_{i\in[u]} k_i'$, we have that:

$$\sum_{i\in[u]} s_i \prod_{j\in[u]} f_{ij}^{k_j} = \sum_{i\in[u]} s_i \prod_{j\in[u]} f_{ij}^{k_j'} \tag{27}$$

Hence, our goal is to prove that (27) implies (26). We formally prove this statement in Lemma 6. Before proving Lemma 6, we will need to prove the following auxiliary lemma.

Lemma 5. *Let $(x_i)_{i=1}^m$,$(y_i)_{i=1}^m$ and $(a_i)_{i=1}^m$ be tuples of positive non-zero reals. If for all positive integers n:*

$$\sum_{i=1}^{m} a_i x_i^n = \sum_{i=1}^{m} a_i y_i^n \tag{28}$$

then the set of entries in $(x_i)_{i=1}^m$ and the set of entries in $(y_i)_{i=1}^m$ are the same.

Proof. Let $\{u_i\}_{i=1}^p$ and $\{v_i\}_{i=1}^q$ be the set of unique entries in $(x_i)_{i=1}^m$ and $(y_i)_{i=1}^m$ respectively. Also, without loss of generality, we may assume an ordering such that $u_1 > u_2 > ... > u_p$ and $v_1 > v_2 > ... > v_q$ and also that $q \geq p$. We can rewrite (28) as:

$$\forall n \in \mathbb{Z}^+ : \sum_{i=1}^{p} c_i u_i^n = \sum_{i=1}^{q} d_i v_i^n \tag{29}$$

As n grows the leading term on LHS is $c_1 u_1^n$ and on the RHS is $d_1 v_1^n$. Hence, it must be :

$$\forall n \in \mathbb{Z}^+ : c_1 u_1^n = d_1 v_1^n$$

Since, u_1, v_1, c_1 and d_1 are non-zero positive reals, we can conclude that $u_1 = v_1$ and $c_1 = d_1$. Hence, we may subtract $c_1 u_1^n$ from both sides in (29) to get :

$$\forall n \in \mathbb{Z}^+ : \sum_{i=2}^{p} c_i u_i^n = \sum_{i=2}^{q} d_i v_i^n \tag{30}$$

We may now repeat the aforementioned argument and infer that $u_2 = v_2$ and $c_2 = d_2$. Furthermore, repeating this argument p times, we can infer that $\{u_i\}_{i=1}^p = \{v_i\}_{i=1}^p$, leaving us with $0 = \sum_{i=q-p+1}^{p} d_i v_i^n$, which is a contradiction as d_i and v_i are positive reals. Hence, we must have that $p = q$. Hence, we have that $\{u_i\}_{i=1}^p$ = $\{v_i\}_{i=1}^q$. Hence, completing the proof. □

[1] For a rigorous proof of why this is true, see Lemma 5 and Lemma 6 in Appendix A.3

Since, $f_{ij} = f_{ji}$, we can see $\{f_{ij}\}$ as a symmetric $u \times u$ matrix (f_{ij}) in $\mathbb{R}^{u\times u}_{>0}$. Hence, the statement that equation (27) implies equation (26) can be formally written as the following Lemma.

Lemma 6. *Let $S = (f_{ij}) \in \mathbb{R}^{u\times u}_{>0}$ be a symmetric matrix and let $(s_i)_{i=1}^{u} \in \mathbb{R}^{u}_{>0}$. If for all $\boldsymbol{k} = \langle k_1, ..., k_u\rangle$ and $\boldsymbol{k'} = \langle k'_1, ..., k'_u\rangle$ such that $k_i, k'_i \in \mathbb{Z}^+$ and $\sum_{i=1}^{u} k_i = \sum_{i=1}^{u} k'_i$, we have that:*

$$\sum_{i=1}^{u} s_i \prod_{j\in[u]} f_{ij}^{k_j} = \sum_{i=1}^{u} s_i \prod_{j\in[u]} f_{ij}^{k'_j} \tag{31}$$

then

$$\forall i, j, i', j' : f_{ij} = f_{i'j'}$$

Proof. Let $\boldsymbol{k}$ be such that $k_p = n$, let $k_i = 0$ for all $i \neq p$. Let $\boldsymbol{k'}$ be such that $k'_q = n$ and $k'_i = 0$ for all $i \neq q$. Then due to (31), we have that:

$$\forall n \in \mathbb{Z}^+ : \sum_{i=1}^{u} s_i(f_{ip})^n = \sum_{i=1}^{u} s_i(f_{iq})^n \tag{32}$$

Hence, due to Lemma 5, we have that the entries in $(f_{ip})_{i=1}^{u}$ and $(f_{iq})_{i=1}^{u}$ form the same set. A similar argument can be repeated for any pair of columns. Hence, all columns in S have the same set of entries, we denote the set of such entries as U.

Let $n = uk$ where $k \in \mathbb{Z}^+$. Let $\boldsymbol{k}$ be such that $k_i = k$ for all $i \in [u]$. Let $\boldsymbol{k'}$ be such that $k'_q = n$ and $k'_i = 0$ for all $i \neq q$. Then due to (31), we have that:

$$\forall k \in \mathbb{Z}^+ : \sum_{i=1}^{u} s_i \prod_{p\in[u]} f_{ip}^{k} = \sum_{i=1}^{u} s_i(f_{iq})^{uk}$$

$$\forall k \in \mathbb{Z}^+ : \sum_{i=1}^{u} s_i(\prod_{p\in[u]} f_{ip})^{k} = \sum_{i=1}^{u} s_i(f_{iq}^{u})^{k}$$

As k grows the leading term on left hand side and right hand side must agree for the equality to hold. Let $c_{i'}(\prod_{p\in[u]} f_{i'p})^k$ and $d_{i''}(f^u_{i''q})^k$ be the leading terms on RHS and LHS respectively. Hence,

$$\forall k \in \mathbb{Z}^+ : c_{i'}(\prod_{p\in[u]} f_{i'p})^k = d_{i''}(f^u_{i''q})^k \tag{33}$$

Using Lemma 5, we have that $\prod_{p\in[u]} f_{i'p} = f^u_{i''q}$. Now, $f_{i''q}$ has to be equal to the maximum term in U, say m. Also, $\prod_{p\in[u]} f_{i'p}$ is a product of all the terms in the p^{th} matrix column of S. Since, each matrix column has the same set of terms U, we have that $\prod_{p\in[u]} f_{i'p} \leq m^u$. But due to (33), we have that, $\prod_{p\in[u]} f_{i'p} = m^u$, which is possible iff:

$$\forall i, j, i', j' : f_{ij} = f_{i'j'}$$

$\square$