



HAL
open science

Limits of Multi-relational Graphs

Juan Aurelio Alvarado, Yuyi Wang, Jan Ramon

► **To cite this version:**

Juan Aurelio Alvarado, Yuyi Wang, Jan Ramon. Limits of Multi-relational Graphs. Machine Learning, 2022. hal-03881631

HAL Id: hal-03881631

<https://inria.hal.science/hal-03881631>

Submitted on 1 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Limits of Multi-relational Graphs

Juan Alvarado · Yuyi Wang · Jan Ramon

Received: date / Accepted: date

Abstract Graphons are limits of large graphs. Motivated by a theoretical problem from statistical relational learning, we develop a generalization of basic results from graphon theory into the “multi-relational” setting. We show that their multi-relational counterparts, which we call multi-relational graphons, are analogically limits of large multi-relational graphs. We extend the cut-distance topology for graphons to multi-relational graphons and prove its compactness and the density of multi-relational graphs in this topology. In turn, compactness enables to prove the large deviation principle for Multi-Relational Graphs (LDP) which enables to prove the most typical random graphs constrained by marginal statistics converge asymptotically to constrained multi-relational graphons with maximum entropy. We show the equivalence between a restricted version of Markov Logic Network and Multi-Relational Graphons with maximum entropy.

Keywords Large multigraph · Graphon Theory · Markov Logic Network · Large Deviation Principle

1 Introduction

Statistical Relational Learning (SRL, [9]) deals with learning probabilistic models from relational data. Many popular SRL frameworks, for instance, Markov Logic Networks (MLNs, Richardson and Domingos [24]), use weighted

✉ J. Alvarado
KU Leuven, Department of Computer Science
E-mail: juan.alvarado@cs.kuleuven.be

Y. Wang
ETH Zurich, Switzerland & CRRIC Zhuzhou Institute. E-mail: yuwang@ethz.ch

J. Ramon
INRIA Lille
E-mail: jan.ramon@inria.fr

logical formulas to encode statistical regularities that hold for the considered problem. Typically, maximum (pseudo-)likelihood estimation is used to compute the weights of the formulas based on training data, which is often a single large example (e.g. a social network). This is problematic because the weights that are learned from this single training example are in general not optimal for examples of different sizes. Jain et al. [12] show a detailed example of a MLN that does not scale on data from different sizes. This turns out to be a fundamental problem, which cannot simply be solved by rescaling the weights. Using statistical terminology, the issue is that models such as MLNs are not projective. Shalizi et al. [26] show that projectivity is essential condition to ensure consistency of estimated probability distribution along samples of different sizes. Jaeger et al. [10] shows some examples of non-projective relational models. An alternative approach to modelling relational structures using MLNs studied in detail in Kuzelka et al. [15] exploits the fact that MLNs can be seen as maximum-entropy models constrained by so called *relational marginal statistics*. Here, the relational marginal statistics are given by first-order logic formulas. In particular, given a first-order logic formula α without quantifiers, the respective relational marginal statistic, denoted $t(\alpha, \omega)$, is the probability that a randomly drawn grounding of the formula α will be satisfied in a given relational structure ω (here, grounding of a formula is a formula obtained by replacing all its variables by constants). For instance, if $\alpha = \text{friends}(x, y)$ then the respective relational marginal statistic “measures” the density of the *friends* relation. Now suppose that we are given formulas $\alpha_1, \alpha_2, \dots, \alpha_m$ and real numbers t_1, t_2, \dots, t_m and we formulate the following optimization problem: find the distribution $P(\omega)$ on possible worlds over some given set of domain constants Δ that (i) satisfies $\mathbb{E}_{\omega \sim P}[t(\alpha_i, \omega)] = t_i$ for each of the formulas $\alpha_1, \dots, \alpha_m$ and (ii) has maximum entropy among all distributions satisfying (i). When ω has infinite number of nodes, Radin et al. [23] show, in the context of Exponential Random Graphs (ERGM), that the most typical worlds ω of MLN, described by one binary and symmetry relation, are the graphons with maximum entropy constrained by $t(\alpha_i, W) = t_i$ where W is a graphon.

Equipped with the view of MLNs as maximum-entropy models constrained by relational marginal statistics, the following problem naturally arises when approximately modelling very large relational structures. Let us fix a set of first order-logic formulas $\alpha_1, \alpha_2, \dots, \alpha_m$ and the values t_1, t_2, \dots, t_m and let us consider the same maximum entropy problem as in the previous paragraph, however, this time we let $|\Delta| \rightarrow \infty$. What can we say about the resulting distribution in the limit? It turns out this is a difficult question and at the moment we do not have much to say about it. However, we can take a look into the theory of *graphons* [16] and see what it can offer for problems like this one.

Graphon theory deals with *limits*, called *graphons*, of large dense graphs. A graphon is any symmetric function $W : [0, 1]^2 \rightarrow [0, 1]$. From a procedural point of view, a graphon can be seen as a probabilistic model for graphs. To sample a random graph $\mathbb{G}(n, W)$ on n vertices from a given graphon W . The

set of vertices of the graph is $\mathbf{V} = \{v_1, v_2, \dots, v_n\}$ and it is obtained by picking n arbitrary objects. The edge set is sampled as follows. Drawing n numbers x_1, x_2, \dots, x_n from $[0, 1]$ uniformly at random. For every possible edge $\{v_{i_1}, v_{i_2}\}$ we uniformly sample a number y_{i_1, i_2} from $[0, 1]$ and if $y_{i_1, i_2} \leq W(x_{i_1}, x_{i_2})$, we add it to the edge set \mathbf{E} . At first this may seem to be a very simple and very restricted model. It may also seem that this model has little to do with the problem we were talking about. However, the opposite is true. Lovasz and Szegedy [17] showed a remarkable result, which can be described using the MLN terminology as follows. Let us consider just one symmetric relation which will represent edges of undirected graphs. Let us fix a set of constant-free and quantifier-free first-order logic formulas, representing some small graphs, e.g. $\alpha_1 = e(x, y)$, $\alpha_2 = e(x, y) \wedge e(y, z)$, $\alpha_3 = e(x, y) \wedge e(y, z) \wedge e(z, x)$. If there is a sequence of undirected graphs of increasing sizes G_1, G_2, G_3, \dots such that the limit $\mathbf{T} = \lim_{i \rightarrow \infty} (t(\alpha_1, G_i), t(\alpha_2, G_i), t(\alpha_3, G_i))$ exists, then there also exists a graphon W such that $\mathbb{E}_{G \sim W} [(t(\alpha_1, G), t(\alpha_2, G), t(\alpha_3, G))] = \mathbf{T}$. In fact, Lovasz and Szegedy showed a stronger result. They showed that graphons are limits of converging sequences of graphs, but to describe this result precisely, we will first need to introduce more background (in particular this result only makes sense after we describe a suitable topology).

Graph homomorphism is the key notion that connects graphons with MLNs. Any constraint based on a graph homomorphism density and a graph density number can be translated into a logical formula with a suitable weight in a MLN. Hence graphons are an alternative for probabilistic modelling of relational structures constrained by relational marginal statistics. The main difference is that graphons maximize Boltzmann Entropy and MLNs maximize Gibbs Entropy. In the parlance of Statistical Mechanics, graphons are microcanonical ensembles of relational structures and MLNs are canonical ensembles. Graphons are also interesting from the computational viewpoint. For instance, unconditional sampling of graphs from a graphon is extremely easy, compared to sampling from MLNs.

A crucial limitation of graphons from the knowledge representation perspective is that they model only probability distributions over simple undirected unlabeled graphs. In this work, we address this limitation and introduce what we will call multi-relational graphons as a direct generalization of graphons into the multi-relational setting. Strikingly, we manage to keep both the simplicity and the theoretical properties of graphons. A multi-relational graphon turns out to be just a vector of graphons, yet all the elegant properties of graphons such as topological compactness carry over to multi-relational graphons. Informally, our main result in the paper is showing the limit of growing random multi-relational graphs constrained by marginal statistics can be obtained by solving an optimization problem in the space of multi-relational graphons.

This potentially opens doors to wider applications of graphons for modelling of large heterogeneous networks: with graphons we could model a network of friends, whereas with multi-relational graphons we can model a network consisting both of friends, acquaintances and families at once. In Section

8.3, we show an example of a MLN that is modeled by multi-relational graphon based on friends and acquaintances relations. Our results may also be potentially useful in the context of the very recently introduced framework called the *AHK model* [10], which is a rather general probabilistic model for relational data. Thus a natural open question would be the generalization of our results about the limits of constrained sequences of random undirected multi-relational graphs to arbitrary-arity relational data. This generalization would give us a principled way to learn AHK models.

This paper is organized as follows. Section 2 introduces Markov logic, multi-relational graphs, and graphon theory. Section 3 introduces multi-relational graphon space, graph homomorphism density, and cut-distance topology. Section 4 proves the compactness of cut-distance topology. Section 5 is devoted to prove that the set of multi-relational graphs is dense in the cut-norm topology. To prove that, we extend the sampling lemmas given by Lovász in his book [16]. Section 6 introduces the large deviation principle for multi-relational Erdős-Rényi random graphs. This principle enables proving that the most typical infinite random graphs constrained by any closed region in the space of multi-relational graphons are solutions of a maximum entropy problem. Section 7 contains the largest proofs of this paper. In this section, we need to introduce yet more mathematical preliminaries oriented to show Alaoglu's Theorem. This theorem states that the unit ball in the dual space of any topological vector space is compact. This theorem enables proving an alternative definition of cut-distance based on couplings. With this alternative definition, the large deviation principle is proven. Section 8 discusses the equivalence of probabilistic models based on multi-relational graphons and a restricted version of MLNs. After that, we show an example of the equivalence between multi-relational graphons and MLNs. In this section, we also show why multi-relational graphon models based on Boltzmann entropy are more natural than maximum entropy models based on Gibbs entropy for modelling multi-relational data. Section 9 describes the relationship to other works with multi-relational graphons. First, we show how multi-relational graphs limits are a special case of Compact Decorated graphs due to Lovász and Szegedy [19]. Second, we show the AHK model is a generalization of multi-relational graphons. Third we discuss the work of constrained graphons with maximum entropy done in the field of Statistical Mechanics. These works can be seen as a special case of constrained multi-graphons with maximum entropy. Finally, Section 10 gives the conclusions of this paper.

2 Background

In this section we cover the necessary background material. Some basic concepts of Topology and Measure Theory are also reviewed in the appendix.

2.1 Basic Notation

We use $\mathbb{1}_X$ to denote the indicator function, i.e., if $x \in X$ then $\mathbb{1}_X(x) = 1$ else $\mathbb{1}_X(x) = 0$. For any positive integer n , we denote by $[n]$ the set of all positive integers smaller than or equal to n . We denote by $\lceil x \rceil$ the least integer greater than or equal to x . For a function $f : X \times X \rightarrow Z$, we denote by f^\top the function which is its transpose, i.e., $\forall x, y \in X : f^\top(x, y) = f(y, x)$. A sequence $x : \mathbb{N} \rightarrow \mathcal{X}$ is denoted by $(x_i)_{i=1}^\infty$.

2.2 First Order Logic

We assume a restricted function-free first-order language defined by a set of constants Δ , a set of variables \mathcal{V} and a set \mathcal{R} of binary predicates. Variables start with lowercase letters and constants start with uppercase letters. A term is a variable or a constant. An atom is $r(a_1, a_2)$ with $a_1, a_2 \in \Delta \cup \mathcal{V}$ and $r \in \mathcal{R}$. A literal is an atom or its negation. An expression is ground if it contains no variables. A substitution is a mapping $\{a_1/t_1, \dots, a_n/t_n\}$ where the a_i are distinct variables and the t_i are terms. If e is an expression and $\theta = \{a_1/t_1, \dots, a_n/t_n\}$ a substitution, then $e\theta$ is the expression e where one has replaced simultaneously the variables a_i with the corresponding terms t_i . A possible world ω is represented as a set of ground atoms that are true in ω . The satisfaction relation \models is defined in the usual way: $\omega \models \alpha$ means that the first-order logic formula α is true in ω . When \mathbf{x} is a list of first-order logic variables then $|\mathbf{x}|$ is used to denote the length of this list.

2.3 Markov Logic Networks

A Markov logic network [24] (MLN) is given as a set of weighted first-order logic formulas (α, w) , where $w \in \mathbb{R}$ and α is a function-free and quantifier-free first-order formula. The semantics are defined w.r.t. the groundings of the first-order formulas, relative to a given finite set of constants Δ , called the domain. An MLN Φ induces the probability distribution over possible worlds $\omega \in \Omega$:

$$p_\Phi(\omega) = \frac{1}{Z} \exp \left(\sum_{(\alpha, w) \in \Phi} w \cdot N(\alpha, \omega) \right),$$

where $N(\alpha, \omega)$ is the number of groundings of α from Δ satisfied in ω , and Z , called the *partition function*, is a normalization constant to ensure that p_Φ is a probability distribution.

In the context of the present paper, we replace $N(\alpha, \omega)$ in the definition of MLNs by the fraction of groundings of α satisfied in ω , i.e.,

$$Q(\alpha, \omega) = \frac{1}{|\Delta|^{|\text{vars}(\alpha)|}} \sum_{\vartheta \in \Theta(\alpha, \Delta)} \mathbb{1}(\omega \models \alpha\vartheta),$$

where $\text{vars}(\alpha)$ is the set of variables occurring in α , $\Theta(\alpha, \Delta)$ is the set of all grounding substitutions of α 's variables using constants from Δ and $\mathbb{1}(\omega \models \alpha\vartheta)$ is the indicator function, which is equal to 1 when $\alpha\vartheta$ is true in the possible world ω , 0 otherwise. Thus, $Q(\alpha, \omega)$ is the fraction of the groundings of α satisfied in ω . With this notation we will write the probability of a possible world $\omega \in \Omega$ as:

$$p_{\Phi}(\omega) = \frac{1}{Z} \exp \left(\sum_{(\alpha, \omega) \in \Phi} w \cdot Q(\alpha, \omega) \right).$$

The reason this representation is more convenient for us is that $Q(\alpha, \omega)$ is closer to the homomorphism densities used in graphon theory, as we discuss in Section 2.5.

Another way to look at Markov logic networks is to view them in the maximum entropy framework as a maximum entropy distribution satisfying the given marginal constraints $\mathbb{E}[Q(\alpha_i, \cdot)] = \theta_i$ (along the lines of [27]). Assuming we are given constraints on the expected values of the statistics $\mathbb{E}[Q(\alpha_i, \cdot)]$ for a set of formulas $\Phi = \{\alpha_1 \dots \alpha_{|\Phi|}\}$, we can define the following max-entropy problem of finding a maximum entropy distribution satisfying the constraints on the expected values.

Relational Marginal Problem (Formulation):

$$\min_{\{P_{\omega} : \omega \in \Omega\}} \sum_{\omega \in \Omega} P_{\omega} \log P_{\omega} \quad s.t. \quad (1)$$

$$\forall i = 1, \dots, m : \sum_{\omega \in \Omega} P_{\omega} \cdot Q(\alpha_i, \omega) = \theta_i \quad (2)$$

$$\forall \omega \in \Omega : P_{\omega} \geq 0, \sum_{\omega \in \Omega} P_{\omega} = 1 \quad (3)$$

Here, the P_{ω} 's are the problem's decision variables, each of which represents the probability of one possible world $\omega \in \Omega$. Line (1) is the maximum Gibbs entropy criterion, which is shown here as the minimization of the negative entropy; Line (2) shows the constraints given by the statistics; and Line (3) provides the normalization constraints for the probability distribution.

Assuming there is a feasible solution satisfying $\forall \omega : P_{\omega} > 0$, the optimal solution of the above maximum entropy problem is an MLN $P_{\omega} = \frac{1}{Z} \exp \left(\sum_{(\alpha_i, w_i) \in \Phi} w_i \cdot Q(\alpha_i, \omega) \right)$ where the parameters $\mathbf{w} = (w_1, \dots, w_m)$ are obtained by maximizing the dual criterion, which also happens to be equivalent to the log-likelihood of the MLN (we refer to [15] for details in the MLN context).

2.4 Graphs and Multi-Relational Graphs

An *undirected graph* \mathbf{G} is a pair (\mathbf{V}, \mathbf{E}) where \mathbf{V} is a finite set of *vertices* and $\mathbf{E} \subseteq \{e \subseteq \mathbf{V} \mid |e| = 2\}$ is a set of *edges*. Two vertices are said to be adjacent (in \mathbf{G}) if they are connected by an edge (of the graph \mathbf{G}). A *multi-relational undirected graph* is a triple $(\mathbf{V}, \mathbf{E}, \lambda)$, where (\mathbf{V}, \mathbf{E}) is an undirected graph and $\lambda : \mathbf{E} \rightarrow 2^\Sigma$ is a function assigning a set of labels from a finite alphabet $\Sigma = \{l_1, \dots, l_r\}$ to every element of \mathbf{E} . Alternatively, a multi-relational graph can be represented as a tuple $(\mathbf{V}, \mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_r)$ where \mathbf{E}_i consists of the edges e for which $l_i \in \lambda(e)$, i.e. which have label l_i among their labels. We will use this latter representation in this paper because it will turn out to be convenient for constructing limits of multi-relational graphs. Note that simple graphs are a special case of multi-relational graphs. Given a multi-relational graph $\mathbf{G} = (\mathbf{V}, \mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_r)$ we use the notation $\mathbf{V}(\mathbf{G}) = \mathbf{V}$ to refer to the set of vertices of \mathbf{G} . The set of multi-relational graphs is denoted by $\mathcal{G}^{[r]}$.

A *homomorphism* from a multi-relational graph $F = (\mathbf{V}^F, \mathbf{E}_1^F, \dots, \mathbf{E}_r^F)$ to a multi-relational graph $G = (\mathbf{V}, \mathbf{E}_1, \dots, \mathbf{E}_r)$ is a mapping $\varphi : \mathbf{V}^F \rightarrow \mathbf{V}$ such that for all $i \in \{1, 2, \dots, r\}$ it holds that if $\{u, v\} \in \mathbf{E}_i^F$ then $\{\varphi(u), \varphi(v)\} \in \mathbf{E}_i$. Informally, homomorphisms are mappings that preserve edges and their labels. The set of all homomorphisms between \mathbf{F} and \mathbf{G} is denoted $\text{Hom}(\mathbf{F}, \mathbf{G})$.

Next we define homomorphism density of \mathbf{F} in \mathbf{G} , denoted $t(\mathbf{F}, \mathbf{G})$, as:

$$t(\mathbf{F}, \mathbf{G}) = \frac{|\text{Hom}(\mathbf{F}, \mathbf{G})|}{|\mathbf{V}(\mathbf{G})|^{|\mathbf{V}(\mathbf{F})|}}. \quad (4)$$

2.5 Graphons

A graphon is the limit of a sequence of growing undirected graphs w.r.t. a certain topology that we describe below. It can be represented as a measurable function $W : [0, 1]^2 \rightarrow [0, 1]$ satisfying $W = W^\top$. The space of graphons \mathcal{W} is defined as

$$\mathcal{W} := \{W : [0, 1]^2 \rightarrow [0, 1] \mid W \text{ is measurable and } W = W^\top\}.$$

The space \mathcal{W} is a subspace of $\mathcal{W}_{\mathbb{R}} = \{W : [0, 1]^2 \rightarrow \mathbb{R} \mid W \text{ is measurable and } W = W^\top\}$. The space $\mathcal{W}_{\mathbb{R}}$ is endowed with the topology induced by the cut-norm,

$$\|W\|_{\square} := \sup_{S, T \subseteq [0, 1]} \left| \int_{S \times T} W(x, y) dx dy \right|.$$

Let Σ be the set of bijective measure preserving maps $\sigma : [0, 1] \rightarrow [0, 1]$. Let $W \in \mathcal{W}$ and let $\sigma \in \Sigma$. Then $W^\sigma : [0, 1]^2 \rightarrow [0, 1]$ is defined by $W^\sigma(x, y) := W(\sigma(x), \sigma(y))$. The space of unlabeled graphons $\widehat{\mathcal{W}}$, shortly called the graphon space, is the quotient of the space of graphons \mathcal{W} w.r.t. the equivalence relation \sim on \mathcal{W} defined by $W \sim V$ iff there is $\sigma \in \Sigma$ such that $W = V^\sigma$. Hence we can also write $\widehat{\mathcal{W}} := \mathcal{W} / \sim$.

The Unlabeled Graphon Space $\widetilde{\mathcal{W}}$ is endowed with a topology induced by the *cut-distance*:

$$\delta_{\square}(W, V) = \inf_{\sigma \in \Sigma} \|W - V^{\sigma}\|_{\square}$$

where $V, W \in \widetilde{\mathcal{W}}$.

Lovász [17] shows that $(\widetilde{\mathcal{W}}, \delta_{\square})$ is a compact topological space. Hence any sequence $(x_n)_{n=1}^{\infty}$ in $\widetilde{\mathcal{W}}$ has a convergent subsequence.

2.5.1 Graphons as Random-Graph Models

Graphons can also be seen as non-parametric random graph models. Given a graphon W and a positive integer n , we can use W to sample random graphs on n vertices as follows:

1. Let $\mathbf{V} = \{v_1, v_2, \dots, v_n\}$.
2. Sample uniformly n numbers x_1, x_2, \dots, x_n from the interval $[0, 1]$.
3. For every v_i, v_j , where $i < j$, sample uniformly a number $y_{i,j}$ from $[0, 1]$.
4. Let $\mathbf{E} = \{\{v_i, v_j\} | i < j \text{ and } y_{i,j} \leq W(x_i, x_j)\}$.
5. Return $\mathbf{G} = (\mathbf{V}, \mathbf{E})$.

Hence, intuitively, we can view a graphon as defining a “probabilistic” adjacency matrix over an uncountable set of vertices.

2.6 Stepfunctions

Stepfunctions are graphons that can be described by a finite number of parameters. First, the probability polytope $P^{(m)}$, where m is a positive integer, is defined by

$$P^{(m)} := \left\{ \pi \in [0, 1]^m \mid \sum_i \pi_i = 1 \right\}.$$

Next we define the notion of m -stepfunction, which will be a central tool in the next sections,

Definition 1 Let λ be the Lebesgue measure on \mathbb{R} , let $P = \{P_1, \dots, P_m\}$ be a partition of $[0, 1]$, let $\pi \in P^{(m)}$ such that $\lambda(P_i) = \pi_i$, let $A \in [0, 1]^{m \times m}$ be a symmetric $m \times m$ matrix on $[0, 1]$. Then an m -stepfunction $(A, \pi) : [0, 1]^2 \rightarrow [0, 1]$ is a function locally constant on $P_i \times P_j$ for all $i, j \in [m]$, more precisely

$$(A, \pi) := \sum_{i,j=1}^m A_{i,j} \mathbb{1}_{P_i \times P_j}.$$

The set of stepfunctions is denoted by $\mathcal{W}^{(m)}$.

3 Multi-Relational Graphons

In this section we introduce *multi-relational graphons*, which, as we show later, are *limits*¹ of convergent series of multi-relational graphs. The multi-relational graphons introduced here can also be seen as special cases of the decorated graphons given in Lovász [19]. However, as we will show, they are simpler to work with.

3.1 Multi-relational Graphon Space

We start by defining *multi-relational graphons*, which are the multi-relational counterparts of graphons. First, let us recall that the space of graphons \mathcal{W} from Section 2.5 is the following

$$\mathcal{W} := \{W : [0, 1]^2 \rightarrow [0, 1] \mid W \text{ is measurable, } W = W^\top\}$$

Since every multi-relational graph G can be represented as a vector of graphs $[G_1, \dots, G_r]$ with a common node set, intuitively we might expect that the limit of a sequence of multi-relational graphs can be represented as a vector of graphons. Later we will see that this definition makes sense and gives us all the desired theoretical properties. Hence, we identify the space of multi-relational graphons, denoted $\mathcal{W}^{[r]}$, with the cartesian product of r copies of the graphon space \mathcal{W} :

Definition 2 (Graphon space)

$$\begin{aligned} \mathcal{W}^{[r]} := \{W : [0, 1]^2 \rightarrow [0, 1]^r \mid W_k : [0, 1]^2 \rightarrow [0, 1] \\ \text{is measurable, } W_k = W_k^\top \text{ for all } k \in [r]\}. \end{aligned}$$

Moreover we extend the cut-norm to the multi-relational graphon space by

$$\|W\|_{\square}^{[r]} = \sum_{i=1}^r \|W_i\|_{\square}. \quad (5)$$

Note that when $r = 1$, $\mathcal{W}^{[r]}$ is reduced to the graphon space \mathcal{W} . Hereafter, we assume that every multi-relational graphon is a vector of r graphons.

Moreover later we need some additional notations for graphon and multi-relational graphon whose values are in $[-1, 1]$.

$$\mathcal{W}_{[-1,1]} := \{W : [0, 1]^2 \rightarrow [-1, 1] \mid W \text{ is measurable and } W = W^\top\}.$$

and

$$\mathcal{W}_{[-1,1]}^{[r]} := \{W : [0, 1]^2 \rightarrow [-1, 1]^r \mid W \text{ is measurable and for all } i \in [r] W_i = W_i^\top\}.$$

¹ What precisely we mean by *being a limit* is explained later.

3.2 Unlabeled Multi-Relational Graphon Space

Next we define the space of unlabeled multi-relational graphons, completely analogously to how the space of unlabeled graphons is defined.

Definition 3 Let $W \in \mathcal{W}^{[r]}$ and let $\sigma \in \Sigma$ then the action of σ on W is defined by,

$$[W_1, \dots, W_r]^\sigma = [W_1^\sigma, \dots, W_r^\sigma].$$

We define the space of unlabeled multi-relational graphons $\widetilde{\mathcal{W}}^{[r]}$ as the quotient of $\mathcal{W}^{[r]}$ by the equivalence relation \sim on $\mathcal{W}^{[r]}$ defined by $W \sim V$ iff there exists an action $\sigma \in \Sigma$ such that $W = V^\sigma$, Hence $\widetilde{\mathcal{W}}^{[r]} := \mathcal{W}^{[r]} / \sim$.

3.2.1 Multi-Relational Graphons as Probabilistic Models

Like graphons, multi-relational graphons can also be seen as non-parametric multi-relational random graph models. Given a multi-relational graphon $W = (W_1, W_2, \dots, W_r)$ and a positive integer n , we can use W to sample random multi-relational graphs on n vertices as follows (we highlight the differences from sampling from simple graphons using text in *italics*):

1. Let $\mathbf{V} = \{v_1, v_2, \dots, v_n\}$.
2. Sample n numbers x_1, x_2, \dots, x_n uniformly from the interval $[0, 1]$.
3. For every v_i, v_j , where $i < j$, *sample r numbers $y_{i,j}^{(1)}, y_{i,j}^{(2)}, \dots, y_{i,j}^{(r)}$ uniformly from $[0, 1]$.*
4. For all $k \in \{1, 2, \dots, r\}$, let $\mathbf{E}_k = \{\{v_i, v_j\} | i < j \text{ and } y_{i,j}^{(k)} \leq W_k(x_i, x_j)\}$.
5. Return $\mathbf{G} = (\mathbf{V}, \mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_r)$.

3.3 Counting Multi-Relational Graph homomorphisms

Let \mathbf{F} and \mathbf{G} be multi-relational graphs with the same number of relations r . Additionally, we assume that \mathbf{G} has node and edges weights. To compute $|Hom(\mathbf{F}, \mathbf{G})|$, we extend the formulas given in Section 5.2.1 in Lovász [16] to count graph homomorphism on weighted graphs to the multi-relational case. For any $k \in [r]$ and vertices $i, j \in V(\mathbf{G})$, let $\alpha_i \in \mathbb{R}_+$ and let $\beta_{k,ij} \in \mathbb{R}_+$ be non-negative weights.

To every map $\phi : V(\mathbf{F}) \rightarrow V(\mathbf{G})$, we assign the weights

$$\alpha_\phi = \prod_{u \in V(\mathbf{F})} \alpha_{\phi(u)}(\mathbf{G})$$

and

$$|Hom_\phi(\mathbf{F}, \mathbf{G})| = \prod_{k=1}^r \prod_{(uv) \in E_k(\mathbf{F})} \beta_{k, \phi(u)\phi(v)}.$$

Moreover, we define

$$|Hom(\mathbf{F}, \mathbf{G})| = \sum_{\phi: V(\mathbf{F}) \rightarrow V(\mathbf{G})} \alpha_\phi |Hom_\phi(\mathbf{F}, \mathbf{G})|. \quad (6)$$

In the sequel of this paper, unless explicitly stated otherwise we will assume that $\alpha_v = 1$ for all $v \in V(\mathbf{G})$. In particular, if we set the node weights $\alpha_\phi = 1$, there follows from Equations (4) and (6) that

$$t(\mathbf{F}, \mathbf{G}) = \sum_{x_1=1}^m \frac{1}{m} \cdots \sum_{x_n=1}^m \frac{1}{m} \prod_{k=1}^r \prod_{uv \in E_k(\mathbf{F})} \beta_{k, x_u x_v} \quad (7)$$

where $m = |V(\mathbf{G})|$ and $n = |V(\mathbf{F})|$. G can be seen as a vector of step-functions on the unit square whose value for the relation k at the square $[(i-1)/|V(G)|, i/|V(G)|] \times [(j-1)/|V(G)|, j/|V(G)|]$ is $\beta_{k, ij}$ for $i, j \in [|V(G)|]$. In this way G is a measurable function, hence when $V(G) \rightarrow \infty$ in a converging sequence we have that $G \rightarrow W$ converges to a measurable function W by Lebesgue's Dominated Convergence Theorem. Hence we arrive to the formula to compute the density of multi-relational graph homomorphisms,

$$t(\mathbf{F}, W) = \int_{[0,1]^{|V(\mathbf{F})|}} \prod_{k=1}^r \prod_{(uv) \in \mathbf{E}_k(F)} W_k(x_u, x_v) \prod_{u \in V(\mathbf{F})} dx_u. \quad (8)$$

Note that the above formula computes the density of multi-relational graph homomorphism from \mathbf{F} into a multi-relational graph W with uncountable many nodes from $[0, 1]$.

In this paper, it is convenient to define F as a generalization of signed graphs from Lovász [16] in the context of multi-relational graphs. Suppose that the edges of each relation E_i of a multi-relational graph F are partitioned into two sets E_i^+ and E_i^- . Then the tuple $F = (V, E_1^+, E_1^-, \dots, E_r^+, E_r^-)$ will be called a multi-relational signed graph. The subgraph density of F in a multi-relational graphon is,

$$t(\mathbf{F}, W) = \int_{[0,1]^{|V(\mathbf{F})|}} \prod_{k=1}^r \prod_{(uv) \in \mathbf{E}_k^+(F)} W_k(x_u, x_v) \prod_{(uv) \in \mathbf{E}_k^-(F)} (1 - W_k(x_u, x_v)) \prod_{u \in V(\mathbf{F})} dx_u. \quad (9)$$

3.4 Cut Distance on $\widetilde{\mathcal{W}}^{[r]}$

We already have a topology based on the cut norm for the space of multi-relational graphons $\mathcal{W}^{[r]}$ that we defined in (5) above. Now, we use it to also endow the space of unlabeled multi-relational graphons $\widetilde{\mathcal{W}}^{[r]}$ with a topology by defining the cut-distance on it.

Let $W, V \in \widetilde{\mathcal{W}}^{[r]}$

$$\delta_{\square}^{[r]}(W, V) = \inf_{\sigma \in \Sigma} \|W - V^\sigma\|_{\square}^{[r]}$$

The following lemma shows bounds for $\delta_{\square}^{[r]}$.

Lemma 1 *Let $W, V \in \widetilde{\mathcal{W}}^{[r]}$ be unlabeled multi-relational graphons, then*

$$\sum_{k=1}^r \delta_{\square}(W_k, V_k) \leq \delta_{\square}^{[r]}(W, V) \leq \|W - V\|_{\square}^{[r]}$$

Proof Note that

$$\begin{aligned} \sum_{k=1}^r \delta_{\square}(W_k, V_k) &= \sum_{k=1}^r \inf_{\sigma_k \in \Sigma} \|W_k - V_k^{\sigma_k}\|_{\square} = \inf_{\sigma_1, \dots, \sigma_r \in \Sigma} \sum_{k=1}^r \|W_k - V_k^{\sigma_k}\|_{\square} \\ &\leq \inf_{\sigma \in \Sigma} \sum_{k=1}^r \|W_k - V_k^{\sigma}\|_{\square} = \delta_{\square}^{[r]}(W, V). \end{aligned}$$

Hence we obtain

$$\sum_{k=1}^r \delta_{\square}(W_k, V_k) \leq \delta_{\square}^{[r]}(W, V) \leq \sum_{k=1}^r \|W_k - V_k\|_{\square}.$$

□

Remark 1 Note that the norm $\|\cdot\|_{\square}^{[r]}$ is the sum of $\|\cdot\|_{\square}$. Hence every result that shows an upper bound for the norm $\|\cdot\|$ on \mathcal{W} for graphons can be extended easily to an upper bound for multi-relational graphons by the sum of individual upper bounds.

4 Compactness on $\widetilde{\mathcal{W}}^{[r]}$

In this section we prove that the space $(\widetilde{\mathcal{W}}^{[r]}, \delta_{\square}^{[r]})$ is compact. To do that we extend the weak regularity lemma for graphons given Corollary 9.13 in Lovász [16].

4.1 Weak Regularity Lemma for multi-relational Graphons

The weak regularity lemma, which is known to hold for graphons, asserts that every graphon can be arbitrarily approximated by stepfunctions. Here we extend the weak regularity lemma to multi-relational graphons.

To state the weak regularity lemma, we need the notion of a *stepping operator*. Intuitively, a stepping operator for multi-relational graphons takes an arbitrary multi-relational graphon and produces a multi-relational graphon which approximates the original graphon and which is a step function.

Definition 4 (Stepping Operator for graphons) Let $P := \{P_1, \dots, P_m\}$ be a partition of $[0, 1]$ and $W \in \mathcal{W}$. Then $T_P(W)$ is the m -stepfunction defined by

$$T_P(W)(x, y) = \frac{1}{\lambda(P_i)\lambda(P_j)} \int_{P_i \times P_j} W(u, v) dudv$$

where $(x, y) \in P_i \times P_j$ and $\lambda(A)$ is the Lebesgue measure of A .

It is clear that the stepping operator T_P is a map from \mathcal{W} to $\mathcal{W}^{(m)}$ where $m = |P|$.

Now we generalize the stepping operator $T_P(W)$ for multi-relational graphons.

Definition 5 (Stepping Operator for Multi-relational graphons) Let $\mathcal{P} = [\mathcal{P}_1, \dots, \mathcal{P}_r]$ be a vector of partitions of $[0, 1]$ and $W \in \mathcal{W}^{[r]}$ be a multi-relational graphon then

$$T_{\mathcal{P}}(W) = [T_{\mathcal{P}_1}(W_1), \dots, T_{\mathcal{P}_r}(W_r)].$$

The next theorem from [16] can be proven using the weak regularity lemma (Lemma 9.9 in Lovász [16]).

Theorem 1 (Corollary 9.13 in Lovász [16]) For every function $W \in \mathcal{W}$ and $k \geq 1$ there is a partition P of $[0, 1]$ into at most k sets with positive measure for which

$$\|W - T_P(W)\|_{\square} \leq \frac{2}{\sqrt{\log_2 k}}$$

Definition 6 Let $P = \{P_1, \dots, P_k\}$ be a partition of $[0, 1]$. We say that P is an equipartition P with k classes if $\lambda(P_i) = 1/k$ for all $i \in [k]$.

Lemma 2 (Lemma 9.15 in Lovász [16]) Let $W \in \mathcal{W}_{[-1,1]}$ and $1 \leq m < k$.

(a) For every m -partition Q of $[0, 1]$ and $k > m$ there is a partition with k classes P refining Q such that

$$\|W - T_P(W)\|_{\square} \leq \frac{2}{\sqrt{\log_2(k/m)}}.$$

(b) For every m -partition Q of $[0, 1]$ and $k > m$ there is an equipartition P with k classes such that

$$\|W - T_P(W)\|_{\square} \leq 2\|W - T_Q(W)\|_{\square} + \frac{2m}{k}$$

Definition 7 (ϵ -partition) Let $\mathcal{P} = \underbrace{[\mathcal{P}_1, \mathcal{P}_1, \dots, \mathcal{P}_1]}_{r \text{ times}}$ be a vector of partitions

of $[0, 1]$ with the same classes. Let $\epsilon > 0$ and let $W \in \mathcal{W}^{[r]}$. We say that \mathcal{P} is an ϵ -partition for W if

$$\|W - T_{\mathcal{P}}(W)\|_{\square}^{[r]} \leq \epsilon.$$

The next theorem proves the existence of an ϵ -partition for any $W \in \mathcal{W}^{[r]}$.

Theorem 2 For every graphon $W \in \mathcal{W}^{[r]}$ and $k \geq 1$ such that $k^{1/r} \in \mathbb{N}$. Then, there exists a partition \mathcal{P} such that

$$\|W - T_{\mathcal{P}}(W)\|_{\square}^{[r]} \leq \frac{2r^{3/2}}{\sqrt{\log_2 k}}.$$

Proof Let $k_1 \geq 1$ be an integer. By Theorem 1, for every $i \in [r]$ there is a k_1 -partition Q_i of $[0, 1]$ such that

$$\|W_i - T_{Q_i}(W_i)\|_{\square} \leq \frac{2}{\sqrt{\log_2 k_1}}.$$

For $s \in [k_1]^r$ let $Q_s^\cap \subseteq [0, 1]$ be the set of all $x \in [0, 1]$ such that for all $i \in [r]$ the point x is in the s_i -th partition class of Q_i . The set $Q^\cap = \{Q_s^\cap \mid s \in [k_1]^r\}$ is a partitioning of $[0, 1]$. Let $k = k_1^r = |Q^\cap|$ be the number of classes of Q^\cap . Q^\cap is a refinement of Q_i for each $i \in [r]$, so there holds

$$\forall i \in [r] : \|W_i - T_{Q^\cap}(W_i)\|_{\square} \leq \frac{2}{\sqrt{\log_2 k_1}} = \frac{2r^{1/2}}{\sqrt{\log_2 k}}$$

and hence

$$\|W - T_{Q^\cap}(W)\|_{\square} \leq \frac{2r^{3/2}}{\sqrt{\log_2 k}}$$

□

4.2 Compactness on $\widetilde{\mathcal{W}}^{[r]}$

We have all ingredients to prove compactness of $\widetilde{\mathcal{W}}^{[r]}$.

Theorem 3 *The space $(\widetilde{\mathcal{W}}^{[r]}, \delta_{\square}^{[r]})$ is compact.*

This proof is almost identical to the proof of Theorem 5.1 in Lovász [18] except that we use a vector of stepfunctions instead of a single stepfunction and we make use of ϵ -partitions.

Proof Let W_1, W_2, \dots be a sequence of vectors of graphons in $\mathcal{W}^{[r]}$. We will construct a subsequence that has a limit in $\widetilde{\mathcal{W}}^{[r]}$.

For each k and n , using Theorem 2, we construct an ϵ -partition $\mathcal{P}_{n,k}$ such that these partitions and the corresponding stepfunctions $W_{n,k} = T_{\mathcal{P}_{n,k}}(W) \in \mathcal{W}^{[r]}$ satisfy the following.

- $\|W_n - W_{n,k}\|_{\square}^{[r]} \leq 1/k$.
- $|\mathcal{P}_{n,k}| = m_k$ (where m_k depends only on k).
- The partition $\mathcal{P}_{n,k+1}$ refines $\mathcal{P}_{n,k}$ for every k .

The rest of the proof follows verbatim from the analogous proof of Theorem 5.1 from [18] except that we use a vector of stepfunctions instead of one stepfunction.

□

Any multi-relational graph $G^{[r]}$ has a natural representation as multi-relational graphon in terms of adjacency matrices:

Definition 8

$$f_k^{G^{[r]}}(x, y) = \begin{cases} 1 & \text{if } (v_{\lceil nx \rceil}, v_{\lceil ny \rceil}) \in E_k(G^{[r]}) \\ 0 & \text{otherwise} \end{cases}$$

where $n = |V(G^{[r]})|$. Hence $f_k^{G^{[r]}}(x, y)$ is a representation of $G^{[r]}$ in $\widetilde{\mathcal{W}}^{[r]}$. From Eqs. (7) and (8), it is not difficult to verify $t(F, G) = t(F, f^G)$.

Compactness of $\widetilde{\mathcal{W}}^{[r]}$ ensures that any sequence of growing multi-relational graphs $(G_n^{[r]})_{n=1}^\infty$ represented by $(f_k^{G_n^{[r]}})_{n=1}^\infty$ has at least a subsequence with a limit. In the next section, we will prove that such limits of sequences of multi-relational graphons $(f_k^{G_n^{[r]}})_{n=1}^\infty$ constrained to a closed set $H \subset \widetilde{\mathcal{W}}^{[r]}$ are solutions of a maximum entropy problem.

5 Density of multi-relational graphs

The weak regularity lemma shows that the set of stepfunctions $\cup_{m>0} \mathcal{W}^{(m)}$ is dense in $\mathcal{W}^{[r]}$ and the set of multi-relational graphs is a subset of $\mathcal{W}^{(m)}$ via the representation $G \mapsto f^G$. To show the space of unlabeled multi-relational graphons is the appropriate limit space for multi-relational graphs, we need to prove that the set of multi-relational graphs $\mathcal{G}^{[r]}$ is dense in $\widetilde{\mathcal{W}}^{[r]}$ via the identification $G \mapsto f^G$. To do that, we extend Theorem 4.7 [4] due to Borgs, Chayes, Lovász, Sós and Vesztergombi to the multi-relational setting. This theorem states that a sample is close to the original graphon with high probability.

We start by extending the notion of W -sampling for multi-relational graphons.

5.1 W -sampling for multi-relational graphons

Definition 9 (Sampling weighted multi-relational graphs) Let $W \in \mathcal{W}^{[r]}$ and let $S = \{x_1, \dots, x_n\} \subset [0, 1]$ be an ordered set, then we denote by $\mathbb{H}(S, W)$ the multi-relational weighted graph $[H_1, \dots, H_r]$ obtained from $W = [W_1, \dots, W_r]$ by assigning the common node set to $V(\mathbb{H}(S, W)) = \{1, \dots, n\}$ and the weight for H_k is equal to

$$\beta_{k,ij}(\mathbb{H}(S, W)) = \begin{cases} W_k(x_i, x_j) & \text{If } i \neq j \\ 0 & \text{If } i = j \end{cases}.$$

for all $1 \leq k \leq r$. If S is a random k -subset of $[0, 1]$ drawn uniformly then we write $\mathbb{H}(n, W)$ instead of $\mathbb{H}(S, W)$.

Definition 10 (Sampling of random graphs from weighted multi-relational graphs) Let H be a weighted multi-relational graph. Then $\mathbb{G}(H)$ is a random multi-relational graph obtained by connecting different nodes i and j by an edge of class k with probability $\beta_{k,ij}(H) \in [0, 1]$ independently of the sampling of other edges. Hence, we can construct a random graph $\mathbb{G}(S, W) = \mathbb{G}(\mathbb{H}(S, W))$.

Definition 11 (Sampling of random graphs from multi-relational graphons) Let $W \in \mathcal{W}^{[r]}$ and let n be a positive integer. Let S be a random k -subset of $[0, 1]$ sampled uniformly. Then $\mathbb{G}(n, W)$ is a random graph obtained by $\mathbb{G}(S, W)$.

Note that the sampling from $\mathbb{G}(n, W)$ extends the sampling procedure of graphs from a graphon given in Section 2.5.1.

Definition 12 (Multi-relational graph parameter) Recall that $\mathcal{G}^{[r]}$ is the set of multi-relational graphs. A multi-relational graph parameter is a function $f : \mathcal{G}^{[r]} \rightarrow \mathbb{R}$ defined on multi-relational, not necessarily loop-free graphs which is invariant under graph isomorphism. For example any multi-relational subgraph density $t(F, \cdot)$ is a multi-relational graph parameter.

Hence it is clear that $f(\mathbb{G}(n, W))$ is a random variable. The next lemma shows a concentration measure inequality for the multi-relational graph parameters.

Lemma 3 (Lemma 4.1 [4]) Let k be a positive integer, and let $C > 0$. Let $Z = (Z_1, \dots, Z_s)$, where Z_1, \dots, Z_s are independent random variables, and Z_i takes values in some measure space (Ω_i, A_i) . Let $f : \Omega_1 \times \dots \times \Omega_s \rightarrow \mathbb{R}$ be a measurable function. Suppose that $|f(x) - f(y)| \leq C$ whenever $x = (x_1, \dots, x_s)$ and $y = (y_1, \dots, y_s)$ differ only in one coordinate. Then

$$\mathbb{P}(f(Z) > E[f(Z)] + \lambda C) < \exp\left(-\frac{\lambda^2}{2s}\right)$$

Remark 2 Let $f : \mathcal{G}^{[r]} \rightarrow \mathbb{R}$ be a multi-relational graph parameter and suppose f is bounded. Let $f_q : [0, 1]^q \rightarrow \mathbb{R}$ with $f_q(S) = f(\mathbb{G}(S, W))$. Then f_q satisfies the condition of the above lemma since there is a C such that $|f_q(x) - f_q(y)| \leq C$ whenever $x = (x_1, \dots, x_q)$ and $y = (y_1, \dots, y_q)$ differ only in one coordinate.

5.1.1 Sampling Lemmas

The goal of this section is the extension of Theorem 4.7 [4] to the multi-relational setting. To do that, we extend a set of lemmas from Chapter 10 in [16] and finally extend Theorem 4.7. We name this extension the second sampling lemma for multi-relational graphons.

Lemma 4 (The First Sampling Lemma for graphons. Lemma 10.6 in [16]) Let $U \in \mathcal{W}_{[-1,1]}$ and let X be a random ordered k -subset of $[0, 1]$. Then with probability at least $1 - 4 \exp(-\sqrt{k}/10)$,

$$-\frac{3}{k} \leq \|U[X]\|_{\square} - \|U\|_{\square} \leq \frac{8}{k^{1/4}}.$$

From the first sampling lemma, we prove the following.

Lemma 5 Let $k > 1$ be a positive integer. Let $U \in \mathcal{W}_{[-1,1]}^{[r]}$ and let X be a random ordered k -subset of $[0, 1]$. Then,

$$\mathbb{E} \left[\|U[X]\|_{\square}^{[r]} - \|U\|_{\square}^{[r]} \right] \leq \frac{14r}{k^{1/4}}.$$

Proof From Remark 1 and the linearity of the expectation, it is sufficient to prove that

$$\mathbb{E}[\|V[X]\|_{\square} - \|V\|_{\square}] \leq \frac{14}{k^{1/4}}.$$

for any $V \in \mathcal{W}_{[-1,1]}$. Let $z = \|V[X]\|_{\square} - \|V\|_{\square}$, thus

$$\begin{aligned} \mathbb{E}[z] &= \mathbb{E}\left[z \mid -\frac{3}{k} \leq z \leq \frac{8}{k^{1/4}}\right] \mathbb{P}\left(-\frac{3}{k} \leq z \leq \frac{8}{k^{1/4}}\right) \\ &\quad + \mathbb{E}\left[z \mid z < -\frac{3}{k} \vee z \geq \frac{8}{k^{1/4}}\right] \mathbb{P}\left(z < -\frac{3}{k} \vee z \geq \frac{8}{k^{1/4}}\right) \end{aligned}$$

In the second term we can bound $\mathbb{E}[z \mid z < -3/k \vee z \geq 8/k^{1/4}] \leq 1$ because $z \in [-1, 1]$, and by applying Lemma 4 on V and $k > 1$ we can bound $\mathbb{P}(z < -3/k \vee z \geq 8/k^{1/4}) \leq 4 \exp(-\sqrt{k}/10)$. In the second term, from Lemma 4 we can bound $\mathbb{E}[z \mid -3/k \leq z \leq 8/k^{1/4}] \leq 8/k^{1/4}$ and the probability is at most 1. There follows

$$E[z] \leq \frac{8}{k^{1/4}} + 4 \exp(-\frac{\sqrt{k}}{10}) \leq \frac{14}{k^{1/4}}.$$

where the last inequality can be verified numerically. \square

We also extend a consequence of Lemma 10.11 [16].

Lemma 6 (Lemma 10.11 [16]) *For every edge-weighted multi-relational graph H with k nodes and edge weights in $[0, 1]$, and for every $\varepsilon \geq 10/\sqrt{k}$, we have*

$$\text{Prob}(\|\mathbb{G}(H) - H\|_{\square} > \varepsilon) \leq \exp(-\varepsilon^2 k^2 / 100).$$

First we prove a consequence of the above lemma in the setting of simple graphs.

Lemma 7 *For every edge-weighted simple graph H with k nodes and edge weights in $[0, 1]$, we have*

$$\mathbb{E}[\|\mathbb{G}(H) - H\|_{\square}] \leq \frac{11}{\sqrt{k}}$$

Proof By applying Lemma 6 with $\varepsilon = 10/\sqrt{k}$, we have

$$\mathbb{P}\left(\|\mathbb{G}(H) - H\|_{\square} > \frac{10}{\sqrt{k}}\right) \leq \exp(-k). \quad (10)$$

We abbreviate $\|\mathbb{G}(H) - H\|_{\square}$ by z . Hence

$$\mathbb{E}[z] = \mathbb{E}\left[z \mid z > \frac{10}{\sqrt{k}}\right] \mathbb{P}\left(z > \frac{10}{\sqrt{k}}\right) + \mathbb{E}\left[z \mid z \leq \frac{10}{\sqrt{k}}\right] \mathbb{P}\left(z \leq \frac{10}{\sqrt{k}}\right)$$

From (10) and knowing the distance $z = \|\mathbb{G}(H) - H\|_{\square}$ is bounded by 1, we have

$$\begin{aligned} \mathbb{E}[z] &\leq \mathbb{E}\left[z \mid z > \frac{10}{\sqrt{k}}\right] \exp(-k) + \mathbb{E}\left[z \mid z \leq \frac{10}{\sqrt{k}}\right] \\ &\leq \exp(-k) + \frac{10}{\sqrt{k}} \leq \frac{11}{\sqrt{k}}. \end{aligned}$$

□

We can extend the above lemma to the multi-relational setting.

Lemma 8 *For every edge-weighted multi-relational graph H with k nodes and edge weights in $[0, 1]$, we have*

$$\mathbb{E}(\|\mathbb{G}(H) - H\|_{\square}^{[r]}) \leq \frac{11r}{\sqrt{k}}$$

Proof From Lemma 7 and Remark 1, the proof is immediate. □

Lemma 9 *Let $V \in \widetilde{\mathcal{W}}^{[r]}$ be a stepfunction of q steps and let k be a positive integer. Then*

$$\mathbb{E}[\delta_{\square}(V, \mathbb{H}(k, V))] \leq r \sqrt{\frac{q}{k}}$$

Proof We follow the same ideas as the proof of Theorem 4.7 in [4].

Let V have steps $J_1, \dots, J_q \subseteq [0, 1]$, and $\lambda(J_i) = \alpha_i$. Let X_1, \dots, X_k be independent random variables that are uniformly distributed on $[0, 1]$, and let Z_i be the number of points X_j that fall into the set J_i . Hence we have

$$\mathbb{E}[Z_i] = \alpha_i k, \quad \text{Var}(Z_i) = (\alpha_i - \alpha_i^2)k \leq \alpha_i k.$$

Construct a partition of $[0, 1]$ into measurable sets J'_1, \dots, J'_q such that $\lambda(J'_i) = Z_i/k$ and

$$\lambda(J'_i \cap J_i) = \min(\alpha_i, Z_i/k).$$

We define a symmetric function $V' \in \mathcal{W}$ such that the value of V' on $J'_i \times J'_j$ is the same as the value of V on $J_i \times J_j$. Then V' is a step function representation of $\mathbb{H}(k, V)$, and it agrees with V on the set $Q = \cup_{i,j=1}^q (J_i \cap J'_i) \times (J_j \cap J'_j)$. Let $\|W\|_1^{[r]} = \sum_i \|W_i\|_1$ be the extension of the L_1 norm to $\mathcal{W}^{[r]}$. Thus we have

$$\begin{aligned} \delta_{\square}(V, \mathbb{H}(k, V)) &\leq \|V - V'\|_{\square}^{[r]} \leq \|V - V'\|_1^{[r]} \leq r(1 - \lambda(Q)) \\ &= r \left(1 - \left(\sum_i \min\left(\alpha_i, \frac{Z_i}{k}\right) \right)^2 \right) \leq 2r \left(1 - \sum_i \min\left(\alpha_i, \frac{Z_i}{k}\right) \right) \\ &= r \sum_i \left| \alpha_i - \frac{Z_i}{k} \right| \leq r \left(q \sum_i \left(\alpha_i - \frac{Z_i}{k} \right)^2 \right)^{1/2}. \end{aligned}$$

Thus we have

$$\delta_{\square}(V, \mathbb{H}(k, V))^2 \leq qr^2 \sum_i \left(\alpha_i - \frac{Z_i}{k} \right)^2.$$

It follows,

$$\begin{aligned} E[\delta_{\square}(V, \mathbb{H}(k, V))^2] &\leq qr^2 \sum_i E \left[\left(\alpha_i - \frac{Z_i}{k} \right)^2 \right] = \frac{qr^2}{k^2} \sum_i \text{Var}(Z_i) \\ &< \frac{qr^2}{k^2} \sum_i k\alpha_i = \frac{qr^2}{k}. \end{aligned}$$

Because $\delta_{\square}(V, \mathbb{H}(k, V)) \geq 0$ there follows by Cauchy-Schwartz

$$E[\delta_{\square}(V, \mathbb{H}(k, V))] \leq r \sqrt{\frac{q}{k}}.$$

□

Now, we have all the ingredients to extend the second sampling lemma.

Lemma 10 *Extension of Theorem 4.7 in [4] (Second Sampling Lemma for Multi-Relational Graphons). Let $k \geq 1$, and let $W \in \widetilde{\mathcal{W}}^{[r]}$ be a multi-relational graphon. Then with probability at least $1 - \exp(-kr/(2 \log_2 k))$,*

$$\delta_{\square}^{[r]}(\mathbb{G}(k, W), W) \leq \frac{36r^{3/2}}{\sqrt{\log_2 k}} \quad (11)$$

One direct consequence of Lemma 10 is that the set of multi-relational graphs is dense in $(\widetilde{\mathcal{W}}^{[r]}, \delta_{\square}^{[r]})$.

Theorem 4 *The set of multi-relational graphs is dense in $(\widetilde{\mathcal{W}}^{[r]}, \delta_{\square}^{[r]})$.*

Proof It is sufficient to prove that for any $W \in \widetilde{\mathcal{W}}^{[r]}$ and $\epsilon > 0$, there is a multi-relational graph G such that $\delta_{\square}^{[r]}(G, W) \leq \epsilon$. There is a sufficient large k such that $\epsilon \geq \frac{37r}{\sqrt{\log_2 k}}$. Then, from Lemma 10 there holds $\delta_{\square}^{[r]}(\mathbb{G}(k, W), W) \leq \epsilon$ with probability at least $1 - \exp(-k/(2 \log_2 k)) > 0$ so at least one of the multi-relational graphs which can be obtained by sampling from $\mathbb{G}(k, W)$ has the required property. □

From the above theorem, any multi-relational graphon can be seen as the limit of a sequence of multi-relational graphs.

6 Large Deviation Principle and the Principle of Maximum-Entropy

The large deviation principle (LDP) establishes lower and upper bounds for the limiting behaviour of a sequence of probability distributions. More precisely [7], let \mathcal{X} be a topological space. A sequence of probability distributions $(\mathbb{P}_n)_{n=1}^{\infty}$ satisfies LDP when there is a function $I : \mathcal{X} \rightarrow \mathbb{R}$, called the rate function such that for every $\Gamma \subset \mathcal{X}$ and sequence $(\epsilon_n)_{n \in \mathbb{N}}$ with limit $\epsilon_n \rightarrow 0$, the probability measure of the set Γ , $\mathbb{P}_n(\Gamma)$, satisfies the following inequality. Recall $\bar{\Gamma}$ and Γ° are closure and interior of $\Gamma \subset \mathcal{X}$.

$$-\inf_{x \in \Gamma^\circ} I(x) \leq \liminf_{n \rightarrow \infty} \epsilon_n \log \mathbb{P}_n(\Gamma) \leq \limsup_{n \rightarrow \infty} \epsilon_n \log \mathbb{P}_n(\Gamma) \leq -\inf_{x \in \bar{\Gamma}} I(x).$$

In this section, we extend the large deviation principle for Erdős-Rényi random graphs due to Chatterjee and Varadhan [5] to multi-relational Erdős-Rényi random graphs.

The negative of the rate function can be interpreted as an entropy function. We will see that $-I(W)$ is a density of entropy of multi-relational random graphs sampled from W . Note that if $\inf_{x \in \Gamma^\circ} I(x) = \inf_{x \in \bar{\Gamma}} I(x)$ then the limiting probability distribution is the probability distribution with maximum entropy. We prove this property for any closed region constrained by densities of multi-relational subgraphs.

Based on the compactness of graphon space, Chatterjee and Varadhan [5] developed the large deviation principle for limiting behavior of the probability measure $\mathbb{P}_{n,p}$ induced by Erdős-Rényi random graphs $G(n,p)$ on the unlabeled graphon space $\widetilde{\mathcal{W}}$.

Here, we first need to extend the Erdős-Rényi random-graph model $G(n,p)$ to multi-relational graphs and define the rate function.

Definition 13 (Erdős-Rényi multi-relational random graphs) Let $p \in [0, 1]$. Then Erdős-Rényi multi-relational random graphs $G^{[r]}(n, p)$ are r -tuple of independent draws of $G(n, p)$, where we identify the vertices across all r graphs.

Definition 14 (Rate function) Let $p \in (0, 1)$. Then the rate function $I_p^{[r]} : \widetilde{\mathcal{W}}^{[r]} \rightarrow \mathbb{R}$ is defined by,

$$I_p^{[r]}(W) = \frac{1}{2} \sum_{k=1}^r \int_{[0,1]^2} I_{0,p}(W_k(x, y)) dx dy$$

where $I_{0,p}(u) = \frac{1}{2}u \log \frac{u}{p} + \frac{1}{2}(1-u) \log \left(\frac{1-u}{1-p} \right)$.

Remark 3 Note that the negative of $I_{0,p}$ is the entropy of the Bernoulli distribution. Hence $-I_p(W)$ can be seen as a density of entropy produced by large random graphs sampled from W . From a statistical mechanics point of view [23], $-I_p(W)$ corresponds to the Boltzmann entropy of the microcanonical ensemble of random edges W .

Let $\mathbb{P}_{n,p}$ be the probability measure on \mathcal{W} induced by $f^{G(n,p)}$ and more generally let $\mathbb{P}_{n,p}^{[r]}$ be the probability measure on $\mathcal{W}^{[r]}$ induced by $f^{G^{[r]}(n,p)}$. In the same way, let $\widetilde{\mathbb{P}}_{n,p}^{[r]}$ be the probability measure on $\widetilde{\mathcal{W}}^{[r]}$ induced by $\widetilde{f}^{G^{[r]}(n,p)}$, where \widetilde{f}^G is the equivalence class of f^G induced by the permutations of rows/columns on f^G .

Now we state our main technical result which is a generalization of the corresponding large deviation result from [5],

Theorem 5 (Extension of Theorem 2.3 [5]) *For each fixed $p \in (0, 1)$, the sequence $\widetilde{\mathbb{P}}_{n,p}^{[r]}$ satisfies a large deviation principle in the cut-distance topology. That is, for every closed set $F \subset \widetilde{\mathcal{W}}$*

$$\limsup_{n \rightarrow \infty} \frac{1}{n^2} \log \widetilde{\mathbb{P}}_{n,p}^{[r]}(F) \leq - \inf_{W \in F} I_p^{[r]}(W) \quad (12)$$

and for any open set $U \subset \widetilde{\mathcal{W}}$,

$$\liminf_{n \rightarrow \infty} \frac{1}{n^2} \log \widetilde{\mathbb{P}}_{n,p}^{[r]}(U) \geq - \inf_{W \in U} I_p^{[r]}(W). \quad (13)$$

6.1 Constraint systems for quantum graphs

In the context of the present paper, it is convenient to define regions of Multi-Relations Graphons by linear combinations of multi-relational graph densities.

Definition 15 (Multi-relational quantum graphs) We say that F is a multi-relational quantum graph if it is the linear combination of a finite number of multi-relational graphs F_i with real coefficients, more precisely

$$F = \sum_i \alpha_i F_i \quad \text{and } \alpha_i \in \mathbb{R}$$

Hence the definition of $t(F, W)$ extends to multi-relational quantum graphs linearly, i.e. $t(F, W) = \sum_i \alpha_i t(F_i, W)$.

Quantum graphs generalize the notion of signed multi-relational graph and offer a convenient way to write succinctly linear combinations of subgraph densities specially when we represent horn clauses using linear combinations of conjunctions.

Let $\mathcal{F} = [\mathcal{F}_1, \dots, \mathcal{F}_k]$ be a finite vector of multi-relational quantum graphs and let $u \in \mathbb{R}_+^k$. We let correspond to (\mathcal{F}, u) a system of constraints $\forall i \in [k] : t(\mathcal{F}_i, W) = u_i$ on the multi-relational quantum graphs $\mathcal{F}_1, \dots, \mathcal{F}_k$, where we call u the density vector. Then we define the feasible region

$$\widetilde{S}(\mathcal{F}, u) = \{W \in \widetilde{\mathcal{W}}^{[r]} \mid t(\mathcal{F}_i, W) = u_i, \text{ for } 1 \leq i \leq |\mathcal{F}|\} \quad (14)$$

We also define an approximate feasible region. In particular, for $\epsilon > 0$,

$$\widetilde{S}(\mathcal{F}, u, \epsilon) = \{W \in \widetilde{\mathcal{W}}^{[r]} \mid |t(\mathcal{F}_i, W) - u_i| \leq \epsilon, \text{ for } 1 \leq i \leq |\mathcal{F}|\} \quad (15)$$

6.2 The most typical multi-relational graph of $\widetilde{S}(\mathcal{F}, u)$

We need an additional ingredient to prove the main result of this section which is the extension of the counting lemma of graphons (Lemma 10.23 in [16]) to multi-relational graphons.

Lemma 11 (Counting lemma for multi-relational graphons) *Let F be a multi-relational graph and let W and V be multi-relational graphons, then*

$$|t(F, W) - t(F, V)| \leq \max_k |E(F_k)| \delta_{\square}^{[r]}(W, V)$$

Thus if F is a quantum multi-relational graph then the map $W \mapsto t(F, W)$ is continuous.

We now define the notion of the most typical multi-relational graphs in $\widetilde{S}(\mathcal{F}, u)$. Let $(G^{[r]}(n, 1/2))_{n=1}^{\infty}$ be a sequence of multi-relational Erdős-Rényi random graphs. From Theorem 4 and the compactness of $\widetilde{\mathcal{W}}^{[r]}$, there is a subsequence $(G^{[r]}(n_i, 1/2))_{i=1}^{\infty}$ and a sequence of positive numbers $(\epsilon_i)_{i=1}^{\infty}$ with $\lim_{i \rightarrow \infty} \epsilon_i = 0$ such that $G^{[r]}(n_i, 1/2) \in \widetilde{S}(\mathcal{F}, u, \epsilon_i)$. By Theorem 4, $\mathcal{G}^{[r]}$ is dense in $\widetilde{\mathcal{W}}^{[r]}$, hence the condition $G^{[r]}(n_i, 1/2) \in \widetilde{S}(\mathcal{F}, u, \epsilon_i)$ holds since $\widetilde{S}(\mathcal{F}, u, \epsilon_i)$ is open for any $\epsilon_i > 0$. Therefore the limit of random multi-relational graphs $G^{[r]}(n_i, 1/2)$ satisfying the constraints $t(\mathcal{F}_i, W) = u \forall i \in [k]$ is defined. We call these limits as the most typical random multi relational graphs

We prove that any typical multi-relational Erdős-Rényi random graph W^* is a solution of $\min_{W \in \widetilde{S}(\mathcal{F}, u)} I_p^{[r]}(W)$, where $p = 1/2$. First we prove the following.

Theorem 6 *Take any $p \in (0, 1)$. Then*

$$\inf_{W \in \widetilde{S}(\mathcal{F}, u)^\circ} I_p^{[r]}(x) = \inf_{W \in \widetilde{S}(\mathcal{F}, u)} I_p^{[r]}(x). \quad (16)$$

and

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \widetilde{\mathbb{P}}_{n,p}^{[r]}(\widetilde{S}(\mathcal{F}, u)) = - \min_{W \in \widetilde{S}(\mathcal{F}, u)} I_p^{[r]}(W).$$

Proof Let $\delta > 0$. Then we define

$$U_\delta = \cap_{i=1}^k \{W \in \widetilde{\mathcal{W}}^{[r]} \mid |t(\mathcal{F}_i, W) - u_i| < \delta\}$$

and

$$V_\delta = \cap_{i=1}^k \{W \in \widetilde{\mathcal{W}}^{[r]} \mid |t(\mathcal{F}_i, W) - u_i| \leq \delta\}$$

From Lemma 11, $t(\mathcal{F}_i, \cdot)$ is continuous under the cut-distance topology, hence U_δ and V_δ are open and closed sets.

Let $\widetilde{\mathbb{P}}_{n,p}^{[r]}(U_\delta)$ and $\widetilde{\mathbb{P}}_{n,p}^{[r]}(V_\delta)$ be the probability measure of U_δ and V_δ . From Theorem 5, it follows that

$$\limsup_{n \rightarrow \infty} \frac{1}{n^2} \log \widetilde{\mathbb{P}}_{n,p}^{[r]}(V_\delta) \leq - \inf_{W \in V_\delta} I_p^{[r]}(W)$$

and

$$\liminf_{n \rightarrow \infty} \frac{1}{n^2} \log \tilde{\mathbb{P}}_{n,p}^{[r]}(U_\delta) \geq - \inf_{W \in U_\delta} I_p^{[r]}(W)$$

and therefore we may write

$$\begin{aligned} - \inf_{W \in U_\delta} I_p^{[r]}(W) &\leq \liminf_{n \rightarrow \infty} \frac{1}{n^2} \log \tilde{\mathbb{P}}_{n,p}^{[r]}(U_\delta) \leq \limsup_{n \rightarrow \infty} \frac{1}{n^2} \log \tilde{\mathbb{P}}_{n,p}^{[r]}(U_\delta) \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{n^2} \log \tilde{\mathbb{P}}_{n,p}^{[r]}(V_\delta) \leq - \inf_{W \in V_\delta} I_p^{[r]}(W) \leq - \inf_{W \in U_{\delta+\delta^2}} I_p^{[r]}(W) \end{aligned}$$

By taking the limit $\delta \rightarrow 0^+$, $\inf_{W \in U_\delta} I_p^{[r]}(W)$ and $\inf_{W \in U_{\delta+\delta^2}} I_p^{[r]}(W)$ are the same and $V_\delta \rightarrow \tilde{S}(\mathcal{F}, u)$. Thus we have,

$$\inf_{W \in \tilde{S}(\mathcal{F}, u)^\circ} I_p^{[r]}(W) = \inf_{W \in \tilde{S}(\mathcal{F}, u)} I_p^{[r]}(W).$$

Since $\tilde{S}(\mathcal{F}, u)$ is compact and Chatterjee and Varadhan proved the function $I_p^{[r]}$ is lower semi-continuous, it follows that we may replace inf by min (c.f. appendix, Theorem 9) and deduce

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \tilde{\mathbb{P}}_{n,p}^{[r]}(\tilde{S}(\mathcal{F}, u)) = - \min_{W \in \tilde{S}(\mathcal{F}, u)} I_p^{[r]}(W).$$

□

Theorem 6 shows that the limiting distribution of the most typical multi-relational Erdős-Rényi random graphs holds the condition

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \tilde{\mathbb{P}}_{n,1/2}^{[r]}(\tilde{S}(\mathcal{F}, u)) = - \min_{W \in \tilde{S}(\mathcal{F}, u)} I_{1/2}^{[r]}(W).$$

The next theorem shows that the limiting distribution of the most typical random graphs is a solution of $\min_{W \in \tilde{S}(\mathcal{F}, u)} I_{1/2}^{[r]}(W)$. To do that, we extend a concentration inequality for growing sequences $(G^{[r]}(n, p))_{n=1}^\infty$ given by Chatterjee and Varadhan, Theorem 3.1 in [5].

Theorem 7 (Concentration inequality for constrained and multi-relational Erdős-Rényi random graphs)

Take any $p \in (0, 1)$ (in our case $p = 1/2$). Let $H \subset \tilde{\mathcal{W}}^{[r]}$ be a compact set such that the condition of Eq. 16 holds. Let H^* be the subset of H where $I_p^{[r]}(\cdot)$ is minimized. Then H^* is non-empty and compact, and for each n , and each $\epsilon > 0$,

$$\text{Prob}(\delta_\square(G^{[r]}(n, p), H^*) \geq \epsilon \mid G^{[r]}(n, p) \in H) \leq \exp(-C(\epsilon, H)n^2)$$

where $C(\epsilon, H)$ is a positive constant depending only on ϵ and H . In particular, if H^* contains only one element h^* , then the conditional distribution of $G^{[r]}(n, p)$ given H converges to the point mass at h^* as $n \rightarrow \infty$.

The original statement of the above theorem includes the condition

$$\inf_{W \in \tilde{S}(\mathcal{F}, u)^\circ} I_p^{[r]}(x) = \inf_{W \in \tilde{S}(\mathcal{F}, u)} I_p^{[r]}(x). \quad (17)$$

which is proven in Theorem 6. The proof of the theorem is omitted since it is identical to the proof of Theorem 3.1 found in Chatterjee and Varadhan [5].

Note that if H is compact the sequence $(G^{[r]}(n, 1/2))_{n=1}^\infty | (H)$ samples uniformly random graphs constrained by H . Hence from the above theorem, we conclude the solutions of $\min_{x \in H} I_{1/2}^{[r]}(x)$ are the most typical infinite multi-relational random graphs constrained to H .

Finally from Theorems 5 and 6, it is straightforward to prove that

Theorem 8 *Let \mathcal{F} be a vector of finite vector of multi-relational quantum graphs $\mathcal{F} = [F_1, \dots, F_k]$ and let $u \in \mathbb{R}^k$ be a vector of multi-relational sub-graph densities. Then the limit $W^*(\mathcal{F}, u)$ of the sequence $(G^{[r]}(n, 1/2))_{n=1}^\infty$ of growing random multi-relational graphs which are uniformly sampled such that $\lim_{n \rightarrow \infty} G^{[r]}(n, 1/2) \in \tilde{S}(\mathcal{F}, u)$ satisfies*

$$W^*(\mathcal{F}, u) = \min_{W \in \tilde{S}(\mathcal{F}, u)} I_p^{[r]}(W) \quad (18)$$

7 Technical Results and Proofs

7.1 Proof of the Second Sampling Lemma

Proof of Lemma 10.

Let us abbreviate $\|W - V\|_\square^{[r]}$ by $d_\square(W, V)$. Recall Theorem 2, for any U there a stepfunction $U_{\mathcal{P}}$ with partition \mathcal{P} of q classes such that,

$$d_\square(U, U_{\mathcal{P}}) \leq \frac{2r^{3/2}}{\sqrt{\log_2 q}}$$

We can also write

$$E[|d_\square(\mathbb{H}(k, U), \mathbb{H}(k, U_{\mathcal{P}}))|] \leq E[|d_\square(\mathbb{H}(k, U), \mathbb{H}(k, U_{\mathcal{P}})) - d_\square(U, U_{\mathcal{P}})|] + d_\square(U, U_{\mathcal{P}})$$

By Lemma 5 and Theorem 2, we have that

$$E[d_\square(\mathbb{H}(k, U), \mathbb{H}(k, U_{\mathcal{P}}))] \leq \frac{14r}{k^{1/4}} + \frac{2r^{3/2}}{\sqrt{\log_2 q}}$$

Thus

$$E[\delta_\square(\mathbb{H}(k, U), \mathbb{H}(k, U_{\mathcal{P}}))] \leq \frac{14r}{k^{1/4}} + \frac{2r^{3/2}}{\sqrt{\log_2 q}}$$

There follows

$$\begin{aligned} E[\delta_{\square}^{[r]}(U, \mathbb{H}(k, U))] &\leq \delta_{\square}^{[r]}(U, U_{\mathcal{P}}) + E[\delta_{\square}^{[r]}(U_{\mathcal{P}}, \mathbb{H}(k, U_{\mathcal{P}}))] + E[\delta_{\square}^{[r]}(\mathbb{H}(k, U_{\mathcal{P}}), \mathbb{H}(k, U))] \\ &\leq \frac{2r^{3/2}}{\sqrt{\log_2 q}} + E[\delta_{\square}^{[r]}(U_{\mathcal{P}}, \mathbb{H}(k, U_{\mathcal{P}}))] + \frac{14r}{k^{1/4}} + \frac{2r^{3/2}}{\sqrt{\log_2 q}} \\ &= \frac{14r}{k^{1/4}} + \frac{4r^{3/2}}{\sqrt{\log_2 q}} + E[\delta_{\square}^{[r]}(U_{\mathcal{P}}, \mathbb{H}(k, U_{\mathcal{P}}))] \end{aligned}$$

Lemma 9 shows an upper bound for $E[\delta_{\square}^{[r]}(U_{\mathcal{P}}, \mathbb{H}(k, U_{\mathcal{P}}))]$, thus

$$E[\delta_{\square}^{[r]}(U, \mathbb{H}(k, U))] \leq \frac{14r}{k^{1/4}} + \frac{4r^{3/2}}{\sqrt{\log_2 q}} + r\sqrt{\frac{q}{k}}.$$

By setting $q = \lceil k^{1/4} \rceil$ we get

$$E[\delta_{\square}^{[r]}(U, \mathbb{H}(k, U))] \leq \frac{14r}{k^{1/4}} + \frac{8r^{3/2}}{\sqrt{\log_2 k}} + r\sqrt{k^{-3/4} + k^{-1}},$$

hence (since $k \geq 1$).

$$E[\delta_{\square}^{[r]}(U, \mathbb{H}(k, U))] \leq \frac{24r^{3/2}}{\sqrt{\log_2 k}}. \quad (19)$$

In order to get an upper bound for $E[\delta_{\square}^{[r]}(U, \mathbb{G}(k, U))]$ we write

$$E[\delta_{\square}^{[r]}(U, \mathbb{G}(k, U))] \leq E[\delta_{\square}^{[r]}(U, \mathbb{H}(k, U))] + E[\delta_{\square}^{[r]}(\mathbb{H}(k, U), \mathbb{G}(k, U))]$$

By Lemma 8 and Eq (19), we have

$$E[\delta_{\square}^{[r]}(U, \mathbb{G}(k, U))] \leq \frac{24r^{3/2}}{\sqrt{\log_2 k}} + \frac{11r}{\sqrt{k}} \leq \frac{35r^{3/2}}{\sqrt{\log_2 k}} \quad (20)$$

Let $f(G) = v(G)\delta_{\square}^{[r]}(G, W)$ be the multi-relational graph parameter. Note that $|f(G) - f(G')| \leq r$ when G and G' differ in only one edge. Hence we can apply Lemma 3 on f , by setting $C = r$, $f(Z) = k\delta_{\square}^{[r]}(\mathbb{G}(k, W), W)$, $E[f(Z)] = k\delta_{\square}^{[r]}(\mathbb{G}(k, W), W)$ and $s = k$.

$$\mathbb{P}(k\delta_{\square}^{[r]}(\mathbb{G}(k, W), W) > kE[\delta_{\square}^{[r]}(\mathbb{G}(k, W), W)] + \lambda r) < \exp(-\frac{\lambda^2}{2k})$$

hence

$$\mathbb{P}(\delta_{\square}^{[r]}(\mathbb{G}(k, W), W) > E[\delta_{\square}^{[r]}(\mathbb{G}(k, W), W)] + \frac{\lambda r}{k}) < \exp(-\frac{\lambda^2}{2k})$$

If $\lambda = \frac{kr^{1/2}}{\sqrt{\log_2 k}}$ then we have

$$\mathbb{P}(\delta_{\square}^{[r]}(\mathbb{G}(k, W), W) \leq E[\delta_{\square}^{[r]}(\mathbb{G}(k, W), W)] + \frac{r^{3/2}}{\sqrt{\log_2 k}}) \geq 1 - \exp(-\frac{kr}{2\log_2 k}).$$

From (20) we know that $E[\delta_{\square}^{[r]}(W, \mathbb{G}(k, W))] \leq \frac{35r^{3/2}}{\sqrt{\log_2 k}}$. Thus there also must hold

$$\mathbb{P}(\delta_{\square}^{[r]}(\mathbb{G}(k, W), W) \leq \frac{35r^{3/2}}{\sqrt{\log_2 k}} + \frac{r^{3/2}}{\sqrt{\log_2 k}}) \geq 1 - \exp\left(-\frac{kr}{2\log_2 k}\right),$$

and hence

$$\mathbb{P}(\delta_{\square}^{[r]}(\mathbb{G}(k, W), W) \leq \frac{36r^{3/2}}{\sqrt{\log_2 k}}) \geq 1 - \exp\left(-\frac{kr}{2\log_2 k}\right).$$

□

7.2 Proof of Counting Lemma for Multi-Relational Graphons

Proof of Lemma 11. This proof follows the same ideas as the proof of Lemma 4.1 in [17]. Let $n = |V(F)|$. Let $E_* = \{(k, i, j) \mid s \in [r] \wedge \{i, j\} \in E_k \wedge i < j\}$ be the collection of all edges of the multi-relational graph. Number the elements in E_* from 1 to $e_* = |E_*|$ and for $s \in E_*$ we denote the s -th triple in E_* by (k_s, i_s, j_s) . We define $X_t : [0, 1]^n \rightarrow \mathbb{R}$ as

$$X_t(x) = \prod_{s=1}^t W_{k_s}(x_{i_s}, x_{j_s}) \prod_{s=t+1}^{e_*} V_{k_s}(x_{i_s}, x_{j_s})$$

Subtracting the integral of two such polynomials for consecutive indices t ,

$$\begin{aligned} & \left| \int_{[0,1]^n} X_t(x) - X_{t-1}(x) \right| \\ & \leq \int_{[0,1]^n} \prod_{s=1}^{t-1} W_{k_s}(x_{i_s}, x_{j_s}) |W_{k_t}(x_{i_t}, x_{j_t}) - V_{k_t}(x_{i_t}, x_{j_t})| \prod_{s=t+1}^{e_*} V_{k_s}(x_{i_s}, x_{j_s}) \\ & \leq \int_{[0,1]^n} |W_{k_t}(x_{i_t}, x_{j_t}) - V_{k_t}(x_{i_t}, x_{j_t})| \\ & \leq \int_0^1 \int_0^1 |W_{k_t}(x_{i_t}, x_{j_t}) - V_{k_t}(x_{i_t}, x_{j_t})| dx_{i_t} dx_{j_t} \\ & \leq \|W_{k_t} - V_{k_t}\|_{\square}. \end{aligned}$$

The last inequality follows from the fact that $\left| \int_{[0,1]^2} W_{k_t}(x, y) - V_{k_t}(x, y) \right|$ is at most the cut norm of $W_{k_t} - V_{k_t}$ (since it uses the rectangle $[0, 1]^2$, rather than $\text{sup}_{S \times T}$). As $t(F, W) = \int_{[0,1]^n} X_{e_*}(x)$ and $t(F, V) = \int_{[0,1]^n} X_0(x)$, we can

use the above inequality to bound

$$\begin{aligned}
|t(F, W) - t(F, V)| &= \left| \sum_{t=1}^{e_*} \int_{[0,1]^n} X_t(x) - X_{t-1}(x) \right| \\
&\leq \sum_{t=1}^{e_*} \left| \int_{[0,1]^n} X_t(x) - X_{t-1}(x) \right| \\
&\leq \sum_{t=1}^{e_*} \|W_{k_t} - V_{k_t}\|_{\square} \\
&= \sum_{k=1}^r |E_k| \|W_k - V_k\|_{\square} \\
&\leq \max_k |E_k| \|W_k - V_k\|_{\square}^r
\end{aligned}$$

Now, noting that $t(F, V) = t(F, V^\sigma)$ for all $\sigma \in \Sigma$, we find

$$|t(F, W) - t(F, V)| \leq \max_k |E_k| \inf_{\sigma \in \Sigma} \|W_k - V_k^\sigma\|_{\square}^r.$$

□

7.3 Proof of the Large Deviation Principle for $G^{[r]}(n, p)$

Here we prove Theorem 5. Our proof technique is inspired by the proof of Theorem 2.3 in Chatterjee and Varadhan [5].

Proof of the upper bound of Theorem 5

Let $\tilde{B}^{[r]}(\tilde{h}, \eta) := \{W \in \tilde{\mathcal{W}}^{[r]} \mid \delta_{\square}^{[r]}(W, \tilde{h}) \leq \eta\}$. First, we claim that to prove the upper bound (12), it is sufficient to prove that for any $\tilde{h} \in \tilde{\mathcal{W}}^{[r]}$,

$$\lim_{\eta \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n^2} \log \tilde{\mathbb{P}}_{n,p}^{[r]}(\tilde{B}^{[r]}(\tilde{h}, \eta)) \leq -I_p^{[r]}(\tilde{h}). \quad (21)$$

Since F is compact, for each $\eta > 0$ we can construct a finite cover $\{\tilde{B}^{[r]}(\tilde{h}_1, \eta), \dots, \tilde{B}^{[r]}(\tilde{h}_k, \eta)\}$ where $\cup_{i=1}^k \tilde{B}^{[r]}(\tilde{h}_i, \eta) \supseteq F$ and $\forall i : \tilde{h}_i \in F$. Thus

$$\tilde{\mathbb{P}}_{n,p}^{[r]}(F) \leq \sum_{i=1}^k \tilde{\mathbb{P}}_{n,p}^{[r]}(\tilde{B}^{[r]}(\tilde{h}_i, \eta)).$$

From (21), there are functions $g_1, g_2 : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ such that

$$\frac{1}{n^2} \log \tilde{\mathbb{P}}_{n,p}^{[r]}(\tilde{B}^{[r]}(\tilde{h}, \eta)) \leq -I_p^{[r]}(\tilde{h}) + g_1(\eta) + g_2(n).$$

with $\lim_{\eta \rightarrow 0} g_1(\eta) = 0$ and $\limsup_{n \rightarrow \infty} g_2(n) = 0$. Hence

$$\tilde{\mathbb{P}}_{n,p}^{[r]}(\tilde{B}^{[r]}(\tilde{h}, \eta) \leq \exp(-n^2(I_p^{[r]}(\tilde{h}) - g_1(\eta) - g_2(n)))$$

for any n and $\eta > 0$. We can thus bound

$$\begin{aligned} \tilde{\mathbb{P}}_{n,p}^{[r]}(F) &\leq \sum_{i=1}^k \tilde{\mathbb{P}}_{n,p}^{[r]}(\tilde{B}^{[r]}(\tilde{h}_i, \eta)) \\ &\leq \sum_{i=1}^k \exp(-n^2(I_p^{[r]}(\tilde{h}_i) - g_1(\eta) - g_2(n))) \\ &\leq k \max_i \exp(-n^2(I_p^{[r]}(\tilde{h}_i) - g_1(\eta) - g_2(n))) \\ &= k \exp\left(-n^2\left(\min_i I_p^{[r]}(\tilde{h}_i) - g_1(\eta) - g_2(n)\right)\right). \end{aligned}$$

Hence

$$\frac{1}{n^2} \log(\tilde{\mathbb{P}}_{n,p}^{[r]}(F)) \leq \min_i I_p^{[r]}(\tilde{h}_i) - g_1(\eta) - g_2(n) + \frac{1}{n^2} \log(k).$$

Note that k is finite for any $\eta > 0$ since F is compact. Then by taking the sup limit $n \rightarrow \infty$ when $\eta > 0$, we get

$$\limsup_{n \rightarrow \infty} \frac{1}{n^2} \log(\tilde{\mathbb{P}}_{n,p}^{[r]}(F)) \leq \min_i I_p^{[r]}(\tilde{h}_i) - g_1(\eta)$$

By taking the limit $\eta \rightarrow 0$, we get

$$\limsup_{n \rightarrow \infty} \frac{1}{n^2} \log(\tilde{\mathbb{P}}_{n,p}^{[r]}(F)) \leq \inf_{\tilde{h} \in F} I_p^{[r]}(\tilde{h}).$$

Thus we have proven that (21) implies (12).

With $B^{[r]}(\tilde{h}, \eta) := \{W \in \mathcal{W}^{[r]} \mid \tilde{W} \in \tilde{B}^{[r]}(\tilde{h}, \eta)\}$, Eq (21) is equivalent to

$$\lim_{\eta \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{P}_{n,p}^{[r]}(B^{[r]}(\tilde{h}, \eta)) \leq -I_p^{[r]}(\tilde{h}). \quad (22)$$

Since $G^{[r]}(n, p)$ has the distribution of an r -tuple of independent $G(n, p)$ random variables, it is not difficult to see that for $\tilde{h} \in \mathcal{W}^{[r]}$ with $\tilde{h} = (\tilde{h}_1 \dots \tilde{h}_r)$,

$$\mathbb{P}_{n,p}^{[r]}(\tilde{h}) = \prod_{k=1}^r \mathbb{P}_{n,p}(\tilde{h}_k)$$

and hence

$$\begin{aligned} \log\left(\mathbb{P}_{n,p}^{[r]}(B^{[r]}(\tilde{h}, \eta))\right) &\leq \log \prod_{k=1}^r \mathbb{P}_{n,p}\left(\text{pr}j_k\left(B^{[r]}(\tilde{h}, \eta)\right)\right) \\ &\leq \log \prod_{k=1}^r \mathbb{P}_{n,p}\left(B(\tilde{h}_k, \eta)\right) \\ &= \sum_{k=1}^r \log \mathbb{P}_{n,p}\left(B(\tilde{h}_k, \eta)\right) \end{aligned}$$

where $prj_k(\cdot)$ is the projection onto the k -th coordinate and $B(\tilde{h}, \eta) = \{W \in \mathcal{W} \mid \delta_{\square}(W, h) \leq \eta\}$. Eq. (11) in the proof of the upper bound of Theorem 2.3 in Chatterjee and Varadhan [5] states that for $\tilde{h} \in \tilde{\mathcal{W}}$,

$$\lim_{\eta \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{P}_{n,p}(B(\tilde{h}, \eta)) \leq -I_p(\tilde{h})$$

Combining these two last inequalities, we get

$$\begin{aligned} & \lim_{\eta \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{P}_{n,p}^{[r]}(B^{[r]}(\tilde{h}, \eta)) \\ & \leq \lim_{\eta \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n^2} \sum_{k=1}^r \log \mathbb{P}_{n,p}(B(\tilde{h}_k, \eta)) \\ & \leq - \sum_{k=1}^r I_p(prj_k(\tilde{h})) \\ & = -I_p^{[r]}(\tilde{h}) \end{aligned}$$

This proves Eq (22) and hence the upper bound.

Proof of the lower bound of Theorem 5

It is sufficient to prove for any $\tilde{h} \in \tilde{\mathcal{W}}^{[r]}$ and arbitrary $\eta > 0$ that

$$\liminf_{n \rightarrow \infty} \frac{1}{n^2} \log \tilde{\mathbb{P}}_{n,p}^{[r]}(\tilde{B}^{[r]}(\tilde{h}, \eta)) \geq -I_p^{[r]}(\tilde{h}). \quad (23)$$

Since $B^{[r]}(\tilde{h}, \eta) \supset \prod_{k=1}^r B(\tilde{h}_k, \eta/r)$, there holds

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n^2} \log \tilde{\mathbb{P}}_{n,p}^{[r]}(\tilde{B}^{[r]}(\tilde{h}, \eta)) &= \liminf_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{P}_{n,p}^{[r]}(B^{[r]}(\tilde{h}, \eta)) \\ &\geq \liminf_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{P}_{n,p}^{[r]} \left(\prod_{k=1}^r B(\tilde{h}_k, \eta/r) \right) \\ &= \sum_{k=1}^r \liminf_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{P}_{n,p} \left(B^{[r]}(\tilde{h}_k, \eta/r) \right) \end{aligned}$$

Applying the lower bound of Theorem 2.3 in Chatterjee and Varadhan [5] we get that for each k ;

$$\liminf_{n \rightarrow \infty} \frac{1}{n^2} \log \mathbb{P}_{n,p} \left(B^{[r]}(\tilde{h}_k, \eta) \right) \geq -I_p(\tilde{h}_k)$$

Combining these inequalities, we get

$$\liminf_{n \rightarrow \infty} \frac{1}{n^2} \log \tilde{\mathbb{P}}_{n,p}^{[r]}(\tilde{B}^{[r]}(\tilde{h}, \eta)) \geq - \sum_{k=1}^r I_p(h_k) = -I_p^{[r]}(h)$$

This proves Eq (23) and hence the lower bound of the theorem.

8 Markov Logic Networks and Multi-Relational Graphons: Two Maximum-Entropy Models

In this section, we discuss the translations between probabilistic models represented in first-order logic and models represented in the multi-relational graph formalism. In particular, we focus on models constrained by relational statistics $Q(\alpha, \omega)$ (the FOL setting) and homomorphism densities (the multi-relational graph setting). As explained in Section 2.2, MLNs are maximum entropy models constrained by expected values of the statistics $Q(\alpha, \omega)$.

8.1 From Multi-Relational Graph Representations to FOL Representations

There is a simple relationship between homomorphism densities $t(\mathbf{F}, \mathbf{G})$ and the statistics $Q(\alpha, \omega)$ that we introduced for Markov logic networks in Section 2.2. First, we show that any multi-relational graph $\mathbf{G} = (\mathbf{V}, \mathbf{E}_1, \dots, \mathbf{E}_m)$ can be trivially represented as a first-order logic possible world ω . This is done as follows. We identify the domain Δ with the set of vertices, i.e. $\Delta = \{v | v \in \mathbf{V}\}$. Then we set

$$\omega_{\mathbf{G}} = \{\lambda_i(v_1, v_2) | v_1, v_2 \in \Delta \text{ and } \{v_1, v_2\} \in \mathbf{E}_i \text{ and } i \in \{1, 2, \dots, m\}\}.$$

The possible world $\omega_{\mathbf{G}}$ plays the same role in the statistic $Q(\alpha, \omega)$ as \mathbf{G} plays in the homomorphism density $t(\mathbf{F}, \mathbf{G})$. It remains to show how to construct a formula α as a counterpart of the graph \mathbf{F} . One way to construct such an α is as follows. Given a graph $\mathbf{F} = (\mathbf{V}^{\mathbf{F}}, \mathbf{E}_1^{\mathbf{F}}, \dots, \mathbf{E}_m^{\mathbf{F}})$, we define α as a conjunction of the following set of first-order logic atoms

$$\mathcal{A} = \{\lambda_i(x_v, x_w) | v, w \in \mathbf{V}^{\mathbf{F}} \text{ and } \{v, w\} \in \mathbf{E}_i^{\mathbf{F}} \text{ and } i \in \{1, 2, \dots, m\}\}.$$

Indeed, it is not difficult to see that if we let $\alpha_{\mathbf{F}} = \bigwedge_{a \in \mathcal{A}} a$, then

$$Q(\alpha_{\mathbf{F}}, \omega_{\mathbf{G}}) = t(\mathbf{F}, \mathbf{G}).$$

Remark 4 For Markov logic networks, it is more common to use formulas in the form of a disjunction of atoms rather than a conjunction. If we construct $\alpha'_{\mathbf{F}} = \bigvee \{-a | a \in \mathcal{A}\}$ then it follows that $Q(\alpha'_{\mathbf{F}}, \omega_{\mathbf{G}}) = 1 - t(\mathbf{F}, \mathbf{G})$. Hence, we can express the same type of constraints using also disjunctions.

It follows from the discussion above that for any set of constraints expressed using homomorphism densities of multi-relational graphs, we can formulate an equivalent set of constraints using first-order logic conjunctions or disjunctions.

8.2 From FOL Representations to Multi-Relational Graph Representations

Next we discuss the other direction. We describe when it is possible to represent constraints in the “world” of relational structures by constraints on multi-relational graphs. One obvious restriction is that we can only represent

constraints on possible worlds where all relations are symmetric, irreflexive and binary. That is we will only be able to model distributions over possible worlds where the following first-order logic sentences are satisfied for all relations r that we use:

$$\begin{aligned} \forall x : \neg r(x, x) \text{ (irreflexivity)} \\ \forall x, y : r(x, y) \Rightarrow r(y, x) \text{ (symmetry)} . \end{aligned}$$

From now on we will assume that all possible worlds ω satisfy these two constraints and we will not always state this explicitly if there is no risk of confusion.

Let α be a conjunction of first-order logic atoms without constants and ω be a possible world on a domain Δ which satisfies the irreflexivity and symmetry constraints. We first show how to construct graphs \mathbf{F}_α and \mathbf{G}_ω such that $Q(\alpha, \omega) = t(\mathbf{F}_\alpha, \mathbf{G}_\omega)$. We define $\mathbf{V}_\mathbf{F} = \{v_x | x \text{ is a FOL variable contained in } \alpha\}$ and $\mathbf{V}_\mathbf{G} = \{v_c | c \in \Delta\}$. Then we define

$$\begin{aligned} \mathbf{F}_\alpha &= (\mathbf{V}^\mathbf{F}, \{\{v_x, v_y\} | \lambda_1(x, y) \in \alpha \text{ or } \lambda_1(y, x) \in \alpha\}, \dots, \\ &\quad \{\{v_x, v_y\} | \lambda_m(x, y) \in \alpha \text{ or } \lambda_m(y, x) \in \alpha\}) \\ \mathbf{G}_\omega &= (\mathbf{V}^\mathbf{G}, \{\{v_x, v_y\} | \lambda_1(x, y) \in \omega \text{ or } \lambda_1(y, x) \in \omega\}, \dots, \\ &\quad \{\{v_x, v_y\} | \lambda_m(x, y) \in \omega \text{ or } \lambda_m(y, x) \in \omega\}) \end{aligned}$$

Next we illustrate the above outlined translation on an example.

Example 1 Let

$$\alpha = \text{friends}(x, y) \wedge \text{friends}(y, z)$$

and

$$\begin{aligned} \omega = \{ &\text{friends}(\text{Alice}, \text{Bob}), \text{friends}(\text{Bob}, \text{Alice}), \text{friends}(\text{Alice}, \text{Eve}), \\ &\text{friends}(\text{Eve}, \text{Alice}), \text{friends}(\text{Bob}, \text{Eve}), \text{friends}(\text{Eve}, \text{Bob})\}. \end{aligned}$$

We now construct the respective graphs \mathbf{F}_α and \mathbf{G}_ω , following the steps outlined above. For \mathbf{F}_α we get

$$\begin{aligned} \mathbf{V}^\mathbf{F} &= \{v_x, v_y, v_z\} \\ \mathbf{E}^\mathbf{F} &= \{\{v_x, v_y\}, \{v_y, v_z\}\}, \end{aligned}$$

and for \mathbf{G}_ω :

$$\begin{aligned} \mathbf{V}^\mathbf{G} &= \{v_{\text{Alice}}, v_{\text{Bob}}, v_{\text{Eve}}\} \\ \mathbf{E}^\mathbf{G} &= \{\{v_{\text{Alice}}, v_{\text{Bob}}\}, \{v_{\text{Alice}}, v_{\text{Eve}}\}, \{v_{\text{Bob}}, v_{\text{Eve}}\}\}. \end{aligned}$$

It is easy to check that $Q(\alpha, \omega) = t(\mathbf{F}, \mathbf{G}) = 12/27$, as expected.

So far we have only explained how to represent constraints which are specified using conjunctions of positive first-order logic literals. It is not difficult to also represent constraints specified by formulas in the form of disjunctions of negative first-order logic literals. If we are given a constraint $\mathbb{E}[Q(\neg a_1 \vee \dots \vee \neg a_m, \cdot)] = t$, where a_1, \dots, a_m are first-order logic atoms, then we can replace it by $\mathbb{E}[Q(a_1 \wedge \dots \wedge a_m, \cdot)] = 1 - t$, which we already know how to represent using multi-relational graphs.

Finally, what remains is to show how to encode constraints represented using either disjunctions or conjunctions of first-order logic literals without the restriction that the literals in them must be all negative or positive, respectively. For this we can exploit the inverse Moebius transform (see e.g. [16, 13, 25], in particular for homomorphism densities, see Eq. 7.11 in Lovász [16]). Let $\alpha = \neg a_1 \wedge \dots \wedge \neg a_l \wedge a_{l+1} \wedge \dots \wedge a_m$ be a conjunction of first-order logic literals. We now need to show how to represent $Q(\alpha, \omega)$ in the graph domain.

Let $\mathbf{F}_\alpha = (\mathbf{V}^{\mathbf{F}}, E_1^+, E_1^-, \dots, E_r^+, E_r^-)$ be a signed multi-relational graph where $\mathbf{V}^{\mathbf{F}}$ is defined above, $E_i^+ = \{\{v_x, v_y\} | \lambda_i(x, y) \in \alpha \text{ or } \lambda_i(y, x) \in \alpha\}$ and $E_i^- = \{\{v_x, v_y\} | \neg \lambda_i(x, y) \in \alpha \text{ or } \neg \lambda_i(y, x) \in \alpha\}$. With this notation we can then write: $Q(\alpha, \omega) = t(\mathbf{F}_\alpha, \mathbf{G}_\omega)$, where t is the subgraph density of signed multi-relational graphs defined in Eq. 9 and \mathbf{G}_ω is defined above.

In conclusion, we have shown that for every constraint that we can write using function-free quantifier-free first-order logic formulas, we can write the equivalent constraint in the graph domain. Note, however, that we restrict ourselves to possible worlds represented over symmetric non-reflexive binary relations.

8.3 An example on how a MLN is represented by Multi-relational Graphons

Using the equivalence between MLN and multi-relational graphs, we transform a MLN Φ into multi-relational graphons with maximum entropy. Φ contains two binary and symmetric relations $friends(x, y)$ and $acquaintances(x, y)$ with the following formulas,

$$\begin{aligned} \infty & : friends(x, y) \Rightarrow friends(y, x) \\ \infty & : acquaintances(x, y) \Rightarrow acquaintances(y, x) \\ \infty & : \neg friends(x, x) \\ \infty & : \neg acquaintances(x, x) \\ w_1 & : friends(x, y) \wedge friends(y, z) \Rightarrow friends(x, z) & (24) \\ w_2 & : friends(x, y) \wedge acquaintances(y, z) \Rightarrow acquaintances(x, z) & (25) \end{aligned}$$

where $w_1, w_2 \in \mathbb{R}$. The probability distribution of the possible worlds ω of Φ is the Gibbs distribution is given by,

$$p_\Phi(\omega) = \frac{1}{Z} \exp(w_1 \cdot Q_1(\omega) + w_2 \cdot Q_2(\omega))$$

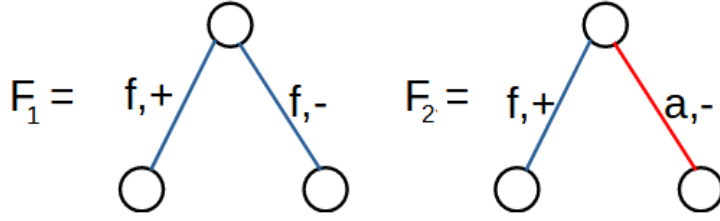


Fig. 1 Two signed multi-relational graphs. The blue edges correspond to 'friends' relations and the red one is 'acquaintances' relation.

Instead, multi-relational graphons enables to formulate a new semantics for Φ by picking the most typical possible worlds of Φ if we assume that $\Delta = [0, 1]$. Hence the possible worlds where the predicates $friends(x, y)$ and $acquaintances(x, y)$ can be satisfied are 2-relational graphons. To pick the most typical possible worlds, we solve the optimization problem.

$$\begin{aligned} & \min_{W \in \overline{\mathcal{W}}^{[r]}} I^{[2]}(W) \text{ subject to} \\ & u_1 = 1 - t(F_1, W) \text{ and } u_2 = 1 - t(F_2, W) \end{aligned}$$

where $\mathbb{E}[Q(\alpha_1, \cdot)] = u_1$ and $\mathbb{E}[Q(\alpha_2, \cdot)] = u_2$ where α_1 is the formula given by Eq. 24 and α_2 is the formula given by Eq. 25, and F_1 and F_2 are signed multi-relational graphs which are shown in Figure 1. Note that the constraints $u_i = 1 - t(F_i, W)$ $i = 1, 2$ come from the Boolean identity $p \Rightarrow q = 1 - (p \wedge \neg q)$.

Once the optimization problem is solved, computing marginal probability of a logical formula ϕ , which is represented by the signed graph F_ϕ , is easy. It is just to compute $\frac{1}{M} \sum_i t(F_\phi, W_i^*)$ where M is the number of global minimizers W_i^* of the above optimization problem.

The key point in the estimation of the most typical worlds from a given set of formulas is to solve the optimization problem in the space of multi-relational graphons. This is a problem posed in a space of infinite many dimensions. Although it is not clear that the problem is computable, prior work on constrained graphons with maximum entropy described in Section 9.3, found by numerical experiments that the global solutions of the maximum entropy problem seem to be stepfunctions.

8.4 Maximum Entropy Distributions

As described in Section 2.3, MLNs are maximum entropy models. In particular they are the model with given statistics $Q(\alpha, \cdot)$, that maximize Gibbs entropy. As we described in detail in the previous sections, constraints equivalent to those on the statistics Q can be also represented in the graph domain using homomorphism densities. If we fix the number of vertices and find the model

that maximizes Gibbs entropy in the graph domain, we end up with a multi-relational variant of *exponential random graph model*² [29]. However, we are not interested in finite graphs and finite relational structures since the large deviation principle shows that graphons maximize the Boltzmann entropy. Let us consider some first-order language with binary relations and denote by Ω_n the set of all possible worlds on a domain of size n , for which the binary relations are symmetric and non-reflexive. It follows from the large deviation principle that if we select from the relational structures Ω_n that violate the given Q -statistics constraints the least and then pick one of them uniformly at random, the limit of this sequence will be a multi-relational graphon that has maximum Boltzmann entropy among the graphons satisfying the constraints that we obtain by transforming the Q -constraints into the graph domain. This is an important result on its own because it justifies the use of maximum-Boltzmann-entropy models for modelling multi-relational data. It could also be potentially relevant as an answer to a question raised in Jaeger and Schulte [10] asking which distributions, satisfying some given constraints one should choose. We answer this question (for undirected multi-relational graphs): one should choose the multi-relational graphons that maximize Boltzmann entropy.

9 Related Work

9.1 Multi-Relational Graphs as Compact Decorated Graphs

Compact Decorated Graphs (*CDG*) are a generalization of simple graphs due to Lovász and Szegedy [19]. *CDG* are defined as follows.

Definition 16 (Compact decorated graphs) Let \mathcal{K} be a compact separable topological space such that \mathcal{K} contains an element $\hat{0}$ representing no edge. A graph decorated with \mathcal{K} is a symmetric function $G : [n] \times [n] \rightarrow \mathcal{K}$. Equivalently, *CDG* is a complete graph K_n with each edge by an element of \mathcal{K} . The set of all \mathcal{F} -decorated graphs with n nodes is denoted by $\mathcal{G}_n(\mathcal{F})$ and $\mathcal{G}(\mathcal{F}) = \cup_{n>0} \mathcal{G}_n(\mathcal{F})$.

Definition 17 (\mathcal{K} -graphon) Let $P(\mathcal{K})$ be the set of Borel probability measures on \mathcal{K} . We denote by $\mathcal{W}(\mathcal{K})$ the set of symmetric functions $W(\mathcal{K}) : [0, 1]^2 \rightarrow P(\mathcal{K})$. The elements of $\mathcal{W}(\mathcal{K})$ are called \mathcal{K} -graphons.

Intuitively, the idea of \mathcal{K} -decorated graph is to model complex relations for every pair of nodes. Hence the element of \mathcal{K} associated with an edge can be seen as a random variable and as the number of nodes goes to infinity, the distribution of the random variable between two nodes $x, y \in [0, 1]$ converges to $W(x, y) \in P(\mathcal{K})$.

² Exponential random graph models are very similar to Markov logic networks.

9.1.1 Graph homomorphism density on $\mathcal{W}(\mathcal{K})$

Graph homomorphism densities on $\mathcal{W}(\mathcal{K})$ may be defined using \mathcal{C} -decorated graphs. Let \mathcal{C} be the set of functions $\mathcal{K} \mapsto \mathbb{R}$. Hence a \mathcal{C} -decorated graph $F = (V, E)$ is a simple graph on which each edge ij is associated with a function $F_{ij} : \mathcal{K} \rightarrow \mathbb{R}$.

Definition 18 Let W be a \mathcal{K} -graphon and let $f \in \mathcal{C}$. Define $W_f : [0, 1]^2 \rightarrow \mathbb{R}$ by $W_f(x, y) = \int_{\mathcal{K}} f dW(x, y)$.

The graph homomorphism density on $\mathcal{W}(\mathcal{K})$ is computed by,

$$t^{CDG}(F, W) = \int_{[0,1]^k} \prod_{1 \leq i < j \leq k} W_{F_{ij}}(x_i, x_j) \prod_{i \in V(F)} dx_i \quad (26)$$

where F is a \mathcal{C} -decorated graph with k nodes and W is a \mathcal{K} -graphon. Note that (26) is generalizing the earlier formula (8) when $r = 1$.

9.1.2 Convergence on CDG

Lovász in [16] shows that any graphon can be seen as the limit of a sequence of growing graphs such that $(G_n)_{n=1}^{\infty}$ converges to $W \in \widetilde{\mathcal{W}}$ iff $(t(F, G_n))_{n=1}^{\infty}$ converges to $t(F, W)$ for every simple graph F . For CDG, there is a similar result. Instead of simple graphs, it uses the notion of Generation system of decorated graphs.

Definition 19 (Generation system. (Definition 2.2 in [19])) We say that a set $\mathcal{F} \subseteq \mathcal{C}$ is dense if for every $\epsilon > 0$ and $f \in \mathcal{C}$ there is a $g \in \mathcal{F}$ such that $|g(x) - f(x)| \leq \epsilon$ for every $x \in \mathcal{K}$. We say that $\mathcal{F} \subseteq \mathcal{C}$ is a generating system if the linear space generated by the elements of \mathcal{F} is dense.

Theorem 9 (Convergence on CDG Theorem 2.6 in [19]) Let \mathcal{F} be a countable generating set and let $(W_n)_{n=1}^{\infty}$ be a sequence of \mathcal{F} -graphons such that $(t(F, W_n))$ is a convergent sequence for every $F \in G(\mathcal{F})$. Then there is a \mathcal{F} -graphon W such that $t(F, W_n) \rightarrow t(F, W)$ for every $F \in G(\mathcal{C})$.

Note that in graphon space the graphs converge to a graphon function W such that $W(x, y) \in [0, 1]$ for all $x, y \in [0, 1]^2$ and in \mathcal{F} -graphon space, the CDG converge to a Borel probability measure on \mathcal{F} instead a number in $[0, 1]$.

9.1.3 Multi-Relational Graphs as Compact Decorated Graphs

Any multi-relational graph can be represented as a compact \mathcal{K} -decorated graph. As follows, Lovász and Szegedy describe a special case of CDG, Parallel Colored Graphs (PCG) where $\mathcal{K} = \{0, 1\}^r$, shown in Example 2.10 [19]. Each edge of a PCG is a general probability distribution on $\{0, 1\}^r$. From Theorem 9, the limits of growing PCG are \mathcal{K} -graphons, $W : [0, 1]^2 \rightarrow [0, 1]^{2^r - 1}$ since a

probability distribution on $\{0, 1\}^r$ can be represented using a vector in $[0, 1]^k$ for $k = 2^r - 1$.

A multi-relational graphon can be seen as a restricted version of the limit of a sequence of *PCGs* in which the probability distribution for every $W(x, y)$ is the product of r independent Bernoulli distributions. Hence the limit object of a convergent sequence of growing multi-relational graphs is a symmetric and measurable function $W : [0, 1]^2 \rightarrow [0, 1]^r$.

Remark 5 The formula for subgraph density on multi-relational graphons (8) can be obtained from the formula for homomorphism density of a *CDG* (26). We use two \mathcal{F} -decorated graphs,

$$\mathcal{F} = \{g_1, g_2 : [0, 1]^r \rightarrow [0, 1] \mid g_1(x) = \prod_k x_k \text{ and } g_2(x) = 1\}$$

where g_1 and g_2 decorate every edge and non-edge. To see that, this yields the subgraph density formula for multi-relational graphs let F^{CDG} be the multi-relational graph F in the form of a \mathcal{F} -graph. By a simple computation we have

$$W_{F_{ij}^{CDG}}(x, y) = \begin{cases} \prod_{k=1}^r W_k(x, y) & \text{if } ij \text{ is connected in } F^{CDG} \\ 1 & \text{otherwise} \end{cases}$$

Hence it is easy to see that $t^{CDG}(F^{CDG}, W) = t(F, W)$.

9.2 The AHK Model

The AHK Model given in Jaeger and Schulte [11] is an infinite exchangeable array which characterizes the probability distributions that are projective on relational structures. A probability distribution is projective when the marginal of the distribution for size- n structures on induced sub-structures of size $k < n$ is equal to the given distribution for size- k structures. More precisely AHK is defined as follows.

Definition 20 Let $[n]_{\neq}^d$ be the set of d -tuples with distinct entries from $[n]^d$. Let $\langle n \rangle^d$ be the set of d -tuples with distinct and ordered entries from $[n]^d$.

Definition 21 Let ω be a relational structure with maximal $\text{arity}(\omega) = a \geq 1$. Let $D_m(\omega)$ $m = 1, \dots, a$ be the arity- m data of ω . $D_m(\omega)$ can be seen as a collection of adjacency matrices referring the connections of m distinct elements. Let $D_m(\omega \downarrow i)$ be the projections of $D_m(\omega)$ in terms of $i \in \langle n \rangle^m$ distinct elements where n is number of elements of ω . Let T_m be the space of possible values of $D_m(\omega \downarrow i)$ when $|i| = m$.

Example 2 Figure 2 shows an example of *AHK* model.

a) The relational structure ω : symmetric and irreflexive relations

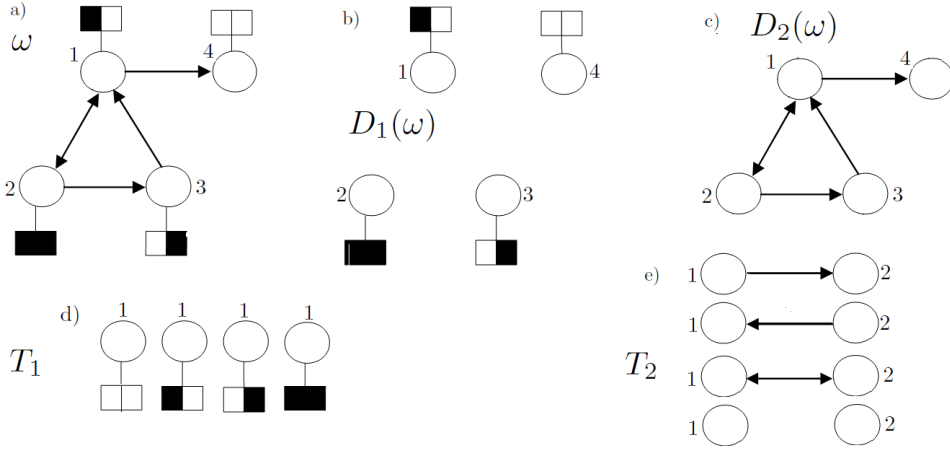


Fig. 2 a) shows a relational structure with 2 unary predicates and 2 binary relations. The circles represent the elements of the domain, boxes attached to the nodes indicate the values of the unary relations, and edges (with arrows) represent the binary relations.

- b) The projection of ω of the unary relation $D_1(\omega)$
- c) The projection of ω of the binary relation $D_2(\omega)$
- d) The possible values for the unary relation T_1
- e) The possible values for the binary relation T_2 .

Definition 22 (AHK model) Let S be a relational signature with maximal $\text{arity}(S) = a \geq 1$. Let $\langle \mathbb{N} \rangle^m$ be the set of tuples from \mathbb{N}^m in which the tuples are distinct and ordered. An AHK model for S is given by

1. A family of i.i.d. latent random variables $\{U_i \mid i \in \langle \mathbb{N} \rangle^m, m = 0, \dots, a\}$, where each U_i is uniformly distributed on $[0, 1]$.
2. A family of random variables $\{D_i \mid i \in \langle \mathbb{N} \rangle^m, m = 0, \dots, a\}$. For $i \in \langle \mathbb{N} \rangle^m$ the variable D_i takes values in T_m .
3. For each $m = 1, \dots, a$ a measurable function $f_m : [0, 1]^{2^m} \rightarrow T_m$ so that
 - For $i = (i_1, \dots, i_m) \in \langle \mathbb{N} \rangle^m$ the value of D_i is defined as $f_m(U_i)$, where

$$U_i = (U_\emptyset, U_{i_1}, \dots, U_{i_m}, U_{(i_1, i_2)}, \dots, U_{(i_{m-1}, i_m)}, \dots, U_{(i_1, \dots, i_m)}) \quad (27)$$

- U_i is a vector containing all $U_{i'}$ -variables with $i' \subset i$ in lexicographic order.
- f_m is invariant under permutations, in the sense that for any permutation $\sigma : [m] \rightarrow [m]$ of $[m]$

$$f_m(U_i^\sigma) = f_m(U_i)^\sigma$$

where U_i^σ is the new multi-dimensional array obtained from U_i where the dimensions i are replaced by the ordered tuple whose entries are $(\sigma(i_1), \dots, \sigma(i_m))$.

9.2.1 Relation between AHK model and Multi-Relational Graphons

Recall that $\mathbb{G}(n, W)$ was mentioned in the Introduction and it is the random graph of n vertices from a given graphon W sampled by picking n numbers x_1, x_2, \dots, x_n from $[0, 1]$ uniformly at random and connecting every possible edge x_i and x_j with probability $W(x_i, x_j)$.

Let $X = \mathbb{G}(n, W)$. The random variable $X_{i_1 i_2}$ can be written as

$$X_{i_1 i_2} = g(U_\emptyset, U_{i_1}, U_{i_2}) \quad (28)$$

where $U_\emptyset, U_{i_1}, U_{i_2} \in [0, 1]$ are uniform random variables and

$$g_W(x_0, x_1, x_2) = \begin{cases} 1 & x_0 \leq W(x_1, x_2) \\ 0 & \text{otherwise} \end{cases}.$$

Note that (28) omits the random variable $U_{(i_1 i_2)}$ so that X is a general 2-dimensional random array. This is because the sampling process of $\mathbb{G}(n, W)$ does not consider the combination of values of i_1 and i_2 to sample $X_{i_1 i_2}$.

For multi-relational random graphs $X = \mathbb{G}^{[r]}(n, W)$ where $W \in \widetilde{\mathcal{W}}^{[r]}$, $X_{i_1 i_2}$ can be represented by the vector $(g_{W_1}(U_\emptyset, U_{i_1}, U_{i_2}), \dots, g_{W_r}(U_\emptyset, U_{i_1}, U_{i_2}))$. Hence it is clear that any multi-relational graphon is an AHK model with maximum arity 2 where the random variable $U_{i_1 i_2}$ in the representation is omitted.

9.3 Prior work on constrained graphons with maximum entropy

Radin et al. [23] proposed graphons with maximum entropy to model large graphs constrained by subgraph density. Radin et al. [22] showed through numerical experiments that graphons with maximum entropy constrained subgraph densities of edges and triangles are stepfunctions.

Kenyon et al. [14] prove rigorously the optimal constrained graphon are stepfunctions when the constraints are on subgraph densities of edge and k -star graphs.

Aristoff et al. [1] prove the large deviation principle for random directed graphs when the constraints are on edge and outward p -star densities.

However it is an open problem to prove that the conjecture is correct. If the conjecture is true, then the maximum entropy graphons can be obtained by solving a finite dimensional non-linear optimization problem.

10 Conclusions

The Large Deviation Principle for Erdős-Rényi multi-relational graphs enables proving the Principle of Maximum Entropy (PME) for multi-relational graphs. This principle states that the most typical random multi-relational graphs constrained by any closed region in the space of multi-relational graphons are

multi-relational graphons with maximum Boltzmann entropy. These Boltzmann entropy maximizers, unlike MLN which are Gibbs entropy maximizers, are a projective model. Thus, multi-relational graphons with maximum entropy are consistent statistical models, in the sense that when the number of nodes of an observed multi-relational network goes to infinity then the network converges to a multi-relational graphon. We show that PME on multi-relational graphs enables the notion of the most typical worlds of a MLN by picking the global solutions of an optimization problem. If Radin's conjecture holds, the optimal computation of the most typical graphons (worlds), constrained by the expectation of logical formulas, can be obtained by a non-linear optimization of finite dimension.

As a candidate theory for MLNs, multi-relational graphon theory has two main issues. The first one is that there is no a rigorous proof of the conjecture raised by Radin et al. in [22], neither for graphons nor multi-relational graphons. However this conjecture has been proved for a special case [14]. The second limitation is that Multi-relational graphon theory is only for symmetric and binary relations. We think that it is possible to extend Multi-relational graphons to higher order relations as long as the relations are symmetric. For non-symmetric relations, the problem is more difficult since it is not clear what is a suitable limiting space for sequences of growing non-symmetric relations [20].

Declarations

Acknowledgements. This work is partially supported by a PhD student scholarship from SENESCYT (Ecuador) and by ERC-StG 240186 MiGraNT: Mining Graphs and Networks, a theory-based approach. The authors would like to thank Ondrej Kuzelka and all the anonymous reviewers for reviewing this work and providing very insightful comments.

References

1. Aristoff, D., Zhu, L.: Asymptotic structure and singularities in constrained directed graphs. *Stochastic Processes and their Applications* **125**(11), 4154–4177 (2015)
2. Aubin, J.P.: *Optima and equilibria: an introduction to nonlinear analysis*, vol. 140. Springer Science & Business Media (1998)
3. Bollobás, B., Riordan, O.: Metrics for sparse graphs. *arXiv preprint arXiv:0708.1919* (2007)
4. Borgs, C., Chayes, J.T., Lovász, L., Sós, V.T., Vesztegombi, K.: Convergent sequences of dense graphs i: Subgraph frequencies, metric properties and testing. *Advances in Mathematics* **219**(6), 1801–1851 (2008)
5. Chatterjee, S., Varadhan, S.: The large deviation principle for the erdős-rényi random graph. *European Journal of Combinatorics* **32**(7), 1000–1017 (2011)
6. Conway, J.B.: *A course in functional analysis*, vol. 96. Springer (2019)
7. Dembo, A., Zeitouni, O.: *Large deviations techniques and applications*, vol. 38. Springer Science & Business Media (2009)
8. Dudley, R.M.: *Real analysis and probability*, vol. 74. Cambridge University Press (2002)

9. Getoor, L., Taskar, B.: Introduction to statistical relational learning, vol. 1. MIT press Cambridge (2007)
10. Jaeger, M., Schulte, O.: Inference, learning, and population size: Projectivity for SRL models. Eighth International Workshop on Statistical Relational AI (StarAI) (2018)
11. Jaeger, M., Schulte, O.: A complete characterization of projectivity for statistical relational models. arXiv preprint arXiv:2004.10984 (2020)
12. Jain, D., Kirchlechner, B., Beetz, M.: Extending Markov logic to model probability distributions in relational domains. In: KI 2007: Advances in Artificial Intelligence, 30th Annual German Conference on AI, KI 2007, pp. 129–143 (2007)
13. Kennes, R., Smets, P.: Computational aspects of the mobius transform. arXiv preprint arXiv:1304.1122 (2013)
14. Kenyon, R., Radin, C., Ren, K., Sadun, L.: Multipodal structure and phase transitions in large constrained graphs. Journal of Statistical Physics **168**(2), 233–258 (2017)
15. Kuželka, O., Wang, Y., Davis, J., Schockaert, S.: Relational marginal problems: Theory and estimation. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18) (2018)
16. Lovász, L.: Large networks and graph limits, vol. 60. American Mathematical Soc. (2012)
17. Lovász, L., Szegedy, B.: Limits of dense graph sequences. Journal of Combinatorial Theory, Series B **96**(6), 933–957 (2006)
18. Lovász, L., Szegedy, B.: Szemerédi’s lemma for the analyst. GAFA Geometric And Functional Analysis **17**(1), 252–270 (2007)
19. Lovász, L., Szegedy, B.: Limits of compact decorated graphs. arXiv preprint arXiv:1010.5155 (2010)
20. Nešetřil, J., de Mendez, P.O.: First-order limits, an analytical perspective. European Journal of Combinatorics **52**, 368–388 (2016)
21. Pedersen, G.K.: Analysis now, vol. 118. Springer Science & Business Media (2012)
22. Radin, C., Ren, K., Sadun, L.: The asymptotics of large constrained graphs. Journal of Physics A: Mathematical and Theoretical **47**(17), 175001 (2014)
23. Radin, C., Sadun, L.: Phase transitions in a complex network. Journal of Physics A: Mathematical and Theoretical **46**(30), 305002 (2013)
24. Richardson, M., Domingos, P.: Markov logic networks. Machine Learning **62**(1-2), 107–136 (2006)
25. Schulte, O., Khosravi, H., Kirkpatrick, A.E., Gao, T., Zhu, Y.: Modelling relational statistics with bayes nets. Machine Learning **94**(1), 105–125 (2014)
26. Shalizi, C.R., Rinaldo, A.: Consistency under sampling of exponential random graph models. Annals of statistics **41**(2), 508 (2013)
27. Wainwright, M.J., Jordan, M.I.: Graphical models, exponential families and variational inference. Foundations and Trends in Machine Learning **1**, 1–305 (2008)
28. Walters, P.: An introduction to ergodic theory, vol. 79. Springer Science & Business Media (2000)
29. Wasserman, S., Pattison, P.: Logit models and logistic regressions for social networks: I. an introduction to markov graphs andp. Psychometrika **61**(3), 401–425 (1996)

Appendix

In this section we provide the necessary concepts from topology, measure theory and compactness on vector spaces. We mostly follow notation and definitions from [8] and [21], except for the definitions of measure preserving map from [28].

Some concepts of topology

Definition 23 (Topological Closure and Interior of a set) Let \mathcal{X} be a topological space and let $A \subset \mathcal{X}$. The topological closure of A and denoted by

\overline{A} is the intersection of all closed sets that contain A . The interior of A and denoted by A° is the union of all open sets contained in A .

Now we define *compact spaces* which are a type of topological spaces.

Definition 24 (Compact space) Let \mathcal{X} be a topological space. We say that \mathcal{X} is compact if for every sequence $(x_i)_{i=1}^\infty$ in \mathcal{X} there is a subsequence $(x_{n_i})_{i=1}^\infty$ that is convergent.

Another alternative definition for compactness is the following. We say that the family of open sets $\{O_i\}$ is an open cover for $A \subset \mathcal{X}$ if $A \subset \cup_i O_i$. Then A is compact if for any cover $\{O_i\}$ of A , it is always possible to pick a subfamily of $\{O_i\}$ such that $A \subset \cup_{j=1}^n O_{i_j}$.

Compactness of a space S ensures the existence of limit points i.e. limits of convergent subsequences of any sequence on S . For instance, every closed and bounded subset of \mathbb{R}^n is compact.

Definition 25 (Lower-continuity) We say that a function f is lower semi-continuous if

$$\liminf_{x \rightarrow x_0} f(x) \geq f(x_0).$$

Lower-continuity is a sufficient condition to state that $\inf_{x \in X} f(x)$ and $\min_{x \in X} f(x)$ are equivalent when X is a compact set.

Theorem 10 (Corollary 1.2 [2]) *If f is a lower-semi continuous and $X \subset \mathcal{X}$ is compact set then $\inf_{x \in X} f(x) = \min_{x \in X} f(x) \in f(X)$.*

Some concepts of measure theory

Next we define σ -algebras, which are building blocks of probability theory.

Definition 26 (σ -Algebra) Let X be a set. We say that a collection of subsets of X , $\mathcal{A}_X \subset \mathcal{P}(X)$, is a σ -algebra on X if (i) \mathcal{A}_X includes \emptyset , (ii) is closed under complement, and (iii) is closed under countable unions.

Having defined σ -algebras, we now define the following notions that we will need in the technical arguments in this paper: *measure functions*, *outer measures* and *measurable sets*, *measure spaces*, *measure functions* and *measure-preserving maps*.

Definition 27 (Measure function) Let \mathcal{A}_X be a σ -algebra on X . We say that a function $\mu : \mathcal{A}_X \rightarrow \mathbb{R}_{\geq 0} \cup \{+\infty\}$ is a measure function if

- $\mu(\emptyset) = 0$
- If $A \in \mathcal{A}$ then $\mu(A) \geq 0$
- For all countable collections of $A_n \in \mathcal{A}$ such that $A_i \cap A_j = \emptyset$ for $i \neq j$ it holds that

$$\mu(\cup_{i=1}^\infty A_i) = \sum_{i=1}^\infty \mu(A_i).$$

Definition 28 (Outer measure) Let \mathcal{A}_X be a σ -algebra and let μ_X be a measure on X . For $E \subset X$ the outer measure of E is defined by,

$$\mu_X^*(E) := \inf \left\{ \sum_{1 \leq n < \infty} \mu(A_n) \mid A_n \in \mathcal{A}_X, E \subseteq \cup_n A_n \right\}.$$

(Here we note that the infimum of an empty set is $+\infty$.)

Next we define *measurable sets* using the just introduced notion of outer-measure.

Definition 29 (Measurable set) We say set $F \subset X$ is a μ^* -measurable, where μ^* is an outer-measure, iff for every set $E \subset X$, we have

$$\mu_X^*(E) = \mu_X^*(E \cap F) + \mu_X^*(E \setminus F)$$

Let $\mathcal{M}_X(\mu_X^*)$ be the collection of sets in X that are measurable. Then $\mathcal{M}_X(\mu^*)$ is σ -algebra and μ^* is a measure function on it (see e.g. Lemma 3.1.8 [8]): hence $(X, \mathcal{M}_X(\mu_X^*), \mu_X^*)$ is a measure space.

Definition 30 (Measure space) We say that the triple $(\mathcal{X}, \mathcal{A}, \mu)$ is a measure space if \mathcal{X} a topological space, \mathcal{A} is the Borel σ -algebra generated by the open sets of \mathcal{X} , and μ is a measure function on \mathcal{A} .

Definition 31 (Measurable function) Let $(\mathcal{X}_i, \mathcal{A}_i, \lambda_i)$, $i = 1, 2$ be measure spaces. Then $f : \mathcal{X}_1 \rightarrow \mathcal{X}_2$ is a measurable function if for any Borel set $A \in \mathcal{A}_2$ then its preimage $f^{-1}(A)$ is a Borel set.

Definition 32 (Measure preserving map) Let $(\mathcal{X}_i, \mathcal{A}_i, \lambda_i)$, $i = 1, 2$ be two measure spaces. Then the measurable function $\sigma : \mathcal{X}_1 \rightarrow \mathcal{X}_2$ is a measure preserving map or Measure preserving transformation if for any Borel set $A \in \mathcal{A}_2$ we have $\lambda_1(\sigma^{-1}(A)) = \lambda_2(A)$.