Markov Random Fields, Homomorphism Counting, and Sidorenko's Conjecture

Péter Csikvári, Nicholas Ruozzi, and Shahab Shams

***Abstract*—Graph covers and the Bethe free energy (BFE) have been useful theoretical tools for producing lower bounds on a variety of counting problems in graphical models, including the permanent and the ferromagnetic Ising model. Here, we investigate weighted homomorphism counting problems over bipartite graphs that are related to a conjecture of Sidorenko. We show that the BFE does yield a lower bound in a variety of natural settings, and when it does yield a lower bound, it necessarily improves upon the lower bound conjectured by Sidorenko. Conversely, we show that there exist bipartite graphs for which the BFE does not yield a lower bound on the homomorphism number. Finally, we use the characterizations developed as part of this work to provide a simple proof of Sidorenko's conjecture in a number of special cases.**

***Index Terms*—Bethe free energy, graph covers, homomorphism counting, Markov random fields, Sidorenko's conjecture**

I. Introduction

A homomorphism from a simple graph $G = (V_G, E_G)$ to a graph $H = (V_H, E_H)$ (possibly with self-loops) is defined to be an adjacency preserving map $h : V_G \to V_H$, i.e., h is a homomorphism if for all $(i,j) \in E_G$ it holds that $(h(i), h(j)) \in E_H$. Our interest here will be in counting the number of homomorphisms from a bipartite graph G into a graph H, denoted as $\hom(G, M^H)$, where $M^H \in \mathbb{R}^{|V_H| \times |V_H|}$ is the unweighted adjacency matrix of graph H such that $M^H_{ij} = 1$ if $(i,j) \in E_H$ and 0 otherwise. Given M^H, the product $\prod_{(i,j)\in E_G} M^H_{h(i),h(j)}$ is equal to one if the mapping h defines a valid homomorphism and zero otherwise. As each map $h : V_G \to V_H$ can be viewed as assigning each vertex in G one of $|V_H|$ labels, we can express the unweighted homomorphism counting problem as

$$\hom(G, M^H) \triangleq \sum_{x \in [|V_H|]^{|V_G|}} \prod_{(i,j)\in E_G} M^H_{x_i,x_j}, \qquad (1)$$

where $[|V_H|] = \{1, \ldots, |V_H|\}$. Note that $\hom(G, H)$ has been used to denote the set of all homomorphisms from G to H, but in this work $\hom(G, M^H)$ denotes the cardinality of set $\hom(G, H)$, i.e., the homomorphism number.

Many combinatorial counting problems can be formulated as homomorphism counting problems from a graph G into a specific graph H [1]. For example, if M^H is the adjacency matrix of a complete graph on n vertices, then $\hom(G, M^H)$ counts the number of ways in which the vertices of G can be colored with n colors such that no two adjacent vertices receive the same color. The problem of counting the number of independent sets in G (subsets of the vertices of G such that no two adjacent vertices are in the set) can be expressed as a homomorphism counting problem by choosing M^H to be the adjacency matrix of a complete graph on two nodes in which exactly one of the nodes has a self loop.

In this work we will focus on a weighted generalization of the homomorphism counting problem to edge weighted graphs, see for example [2], [3]. In the weighted setting, each edge of the target graph H is associated with a nonnegative weight. Let G be a simple graph and $M \in \mathbb{R}^{|V_H|\times|V_H|}_{\geq 0}$ a symmetric nonnegative matrix such that M_{ij} is the weight associated with the edge (i,j) in the target graph. A zero weight corresponds to the absence of an edge in the target graph, so in general, we can assume that the target graph is a complete graph with self-loops. The weight of a map h from the vertices of the bipartite graph G into the target graph H is then defined to be

$$\prod_{(i,j)\in E_G} M_{h(i),h(j)},$$

and the weighted homomorphism counting problem is defined as

$$\hom(G, M) \triangleq \sum_{x \in [|V_H|]^{|V_G|}} \prod_{(i,j)\in E_G} M_{x_i,x_j}. \qquad (2)$$

This paper was presented in part at the IEEE International Symposium on Information Theory (ISIT), July 2019.

This work was supported by the Marie Skłodowska-Curie Individual Fellowship grant no. 747430 and the Counting in Sparse Graphs Lendület Research Group of Alfréd Rényi Institute of Mathematics, the DARPA Explainable Artificial Intelligence (XAI) program under contract number N66001-17-2-4032, and National Science Foundation grant III-1527312.

Péter Csikvári is at the Alfréd Rényi Institute of Mathematics, Budapest, Hungary (peter.csikvari@gmail.com) Nicholas Ruozzi and Shahab Shams are part of the Department of Computer Science at the University of Texas at Dallas, Richardson, TX 75080 (nicholas.ruozzi@utdallas.edu and shahb.shams@utdallas.edu).

The above discussion generalizes to directed graphs in a straightforward way, i.e., maps from G to H that preserve directed edges. In this work, we will specifically be interested in the case in which $G = (A_G, B_G, E_G)$ is a directed bipartite graph in which all edges are directed from A_G to B_G, and H is a weighted directed graph whose edge weights are given by a nonnegative matrix $M \in \mathbb{R}_{\geq 0}^{|V_H|\times|V_H|}$. The weighted homomorphism counting problem in (directed) bipartite graphs is then to compute

$$\hom(G, M) = \sum_{x\in[|V_H|]^{|V_G|}} \prod_{i\in A_G} \prod_{j\in N_G(i)} M_{x_i,x_j}, \quad (3)$$

where $V_G = A_G \cup B_G$ and $N_G(i) \subseteq B_G$ is the set of neighbors of the vertex $i \in A_G$. Bipartite graphs are of particular interest as Simonovits [4] and Sidorenko [5], [6] conjectured a lower bound, widely known as Sidorenko's conjecture, on $\hom(G, M)$ in this case.

Conjecture I.1 (Sidorenko's Conjecture)**.** For a bipartite graph $G = (A_G, B_G, E_G)$ and an arbitrary weight matrix $M \in \mathbb{R}_{\geq 0}^{m\times m}$,

$$\begin{aligned}\hom(G, M) &\geq \hom(K_2, M)^{|E_G|} m^{|V_G|-2|E_G|} \\ &\triangleq \hom_{\mathrm{S}}(G, M), \end{aligned} \quad (4)$$

where K_2 is the complete graph on two vertices. Note that since $M \in \mathbb{R}_{\geq 0}^{m\times m}$, $\hom(K_2, M) = \sum_{x_1,x_2\in[m]} M_{x_1,x_2}$, equals to the sum over all the entries of matrix M. In the special case that M^H corresponds to the adjacency matrix of an unweighted graph H, the lower bound given by Sidorenko's conjecture is equivalent to

$$\hom(G, M^H) \geq |V_H|^{|V_G|} \left(\frac{2|E_H|}{|V_H|^2}\right)^{|E_G|}.$$

Different approaches have led to a partial resolution of this conjecture in special cases: the case in which G is a tree or an even cycle [4], [6], a cube [7], or any bipartite graph that has one vertex complete to the other side [8], [9]. Many of these results can be proved using information theoretic inequalities and clever repeated application of Jensen's inequality [10], [11]. Separately, work on approximate counting in Markov random fields based on graph covers and the Bethe free energy [12] has produced lower bounds for a variety of counting problems: matrix permanents [13], real-stable polynomials [14], the ferromagnetic Ising model with arbitrary external field [15], the ferromagnetic Potts model with uniform external field [15], weight enumerators of linear codes [15], the weighted homomorphism counting problem when $\mathrm{rank}(M) \leq 2$ [15], and others. In particular, this line of work yields a proof of Conjecture I.1 in the cases that $M \in \mathbb{R}_{\geq 0}^{2\times 2}$ [16] or $M = aa^T + bb^T$ for $a, b \in \mathbb{R}_{\geq 0}^{m}$ [15].

Given the positive results for homomorphism counting in the above special cases, a natural question is whether or not a similar result can be shown for the general weighted homomorphism counting problem over bipartite graphs. In this work, we explore this question using, again, the Bethe free energy optimization problem from statistical physics (see Section II for the definition), which defines an approximation to $\hom$ that is, in general, neither an upper nor a lower bound. Denoting the Bethe free energy approximation as $\hom_{\mathrm{B}}$, we will be interested in the veracity of the following conjecture.

Conjecture I.2. If G is a bipartite graph, then

$$\hom(G, M) \geq \hom_{\mathrm{B}}(G, M).$$

This paper provides both positive and negative results related to Conjecture I.2. Section III proves that for all bipartite graphs G and all $M \in \mathbb{R}_{\geq 0}^{m\times m}$, $\hom_{\mathrm{B}}(G, M) \geq \hom_{\mathrm{S}}(G, M)$. Consequently, proving Conjecture I.2 would resolve Sidorenko's conjecture in the affirmative. On the other hand, Section IV disproves Conjecture I.2 by constructing an explicit counterexample, a bipartite graph G and a matrix $M \in \mathbb{R}_{\geq 0}^{m\times m}$, for which $\hom(G, M) < \hom_{\mathrm{B}}(G, M)$. Section V provides two simpler formulations of Conjecture I.2. Those two formulations are then used to demonstrate that $\hom_{\mathrm{B}}$ yields a better lower bound than Sidorenko's bound in a number of interesting special cases. Finally, we conclude with a few results on an interesting special case of the conjecture that has not received much attention, i.e., the case of doubly stochastic matrices.

II. Preliminaries

In this section, we review Markov random fields (MRFs), the Bethe free energy, log-supermodularity, and related results that are relevant to this work.

A. *Markov random fields and partition functions*

A pairwise Markov random field is defined by a graph $G = (V_G, E_G)$ together with a collection of nonnegative potential functions $\phi_i : \mathcal{X} \to \mathbb{R}_{\geq 0}$ for each $i \in V_G$ and $\psi_{ij} : \mathcal{X}^2 \to \mathbb{R}_{\geq 0}$ for each $(i, j) \in E_G$. The product of these potential functions defines a probability distribution

$$p(X) = \frac{1}{Z(G; \phi, \psi)} \prod_{i\in V_G} \phi_i(X_i) \prod_{(i,j)\in E_G} \psi_{ij}(X_i, X_j),$$

where $X \in \mathcal{X}^{|V_G|}$ is a vector of random variables, $\mathcal{X}$ is the domain of the random variables, e.g., $\mathcal{X} = [m]$ or $\mathbb{R}$, and Z is the normalizing constant/partition function defined as

$$Z(G; \phi, \psi) \triangleq \sum_{X\in\mathcal{X}^{|V_G|}} \prod_{i\in V_G} \phi_i(X_i) \prod_{(i,j)\in E_G} \psi_{ij}(X_i, X_j).$$

A detailed review of Markov random fields and their properties can be found in [17].

Given a bipartite graph $G = (A_G, B_G, E_G)$ and $M \in \mathbb{R}^{m\times m}_{\geq 0}$, $\hom(G, M)$ is the partition function of an MRF over the state space $\mathcal{X} = [m]$, where for all $i \in A_G$ and $j \in N_G(i)$ the potential function on the edge (i, j) is given by

$$\psi_{ij}(x_i, x_j) = M_{x_i,x_j}.$$

Note that $\phi_i(x_i) = 1$ for all $i \in A_G \cup B_G$.

A variety of different approximations to the partition function of an MRF have been proposed. One particular approximation, the Bethe free energy approximation [18], has been shown to provide lower bounds on the partition function, Z, for certain nice families of MRFs, e.g., log-supermodular models [16]. We consider this approximation here in the context of the MRF for the weighted homomorphism counting problem with weight matrix $M \in \mathbb{R}^{m\times m}_{\geq 0}$.

Let the *local marginal polytope*, $\mathcal{T}$, consist of vectors of probability distributions. There is exactly one entry in the vector $\tau \in \mathcal{T}$ for each $i \in V_G$ and each edge $(i, j) \in E_G$. The marginals in any given vector should agree on single variable overlaps. More formally, $\mathcal{T}(G)$ consists of all vectors of probability distributions τ such that

$$\forall (i,j) \in E_G, x_i \in [m], \sum_{x_j} \tau_{ij}(x_i, x_j) = \tau_i(x_i),$$
$$\forall i \in V_G, \sum_{x_i} \tau_i(x_i) = 1,$$
$$\forall i \in V_G, x_i \in [m], \tau_i(x_i) \geq 0,$$
$$\forall (i,j) \in E_G, x_i \in [m], x_j \in [m], \tau_{ij}(x_i, x_j) \geq 0.$$

The Bethe free energy approximation is given by

$$\log F_{\mathrm{B}}(G, \tau; M) = U(G, \tau; M) + \widetilde{E}(G, \tau),$$

where U is the negative energy,

$$U(G, \tau; M) \triangleq \sum_{i\in A_G} \sum_{j\in N_G(i)} \sum_{x_i,x_j} \tau_{ij}(x_i, x_j) \log M_{x_i,x_j},$$

and $\widetilde{E}$ is an entropy approximation,

$$\widetilde{E}(G, \tau) \triangleq -\sum_{i\in V_G} \sum_{x_i} \tau_i(x_i) \log \tau_i(x_i) - \sum_{(i,j)\in E_G} \sum_{x_i,x_j} \tau_{ij}(x_i, x_j) \log \frac{\tau_{ij}(x_i, x_j)}{\tau_i(x_i)\tau_j(x_j)}.$$

The Bethe partition function for the weighted homomorphism counting problem is obtained by exponentiating the maximum value achieved by this approximation over $\mathcal{T}(G)$. More specifically,

$$\hom_{\mathrm{B}}(G, M^H) \triangleq \exp\left(\max_{\tau\in\mathcal{T}(G)} F_{\mathrm{B}}(G, \tau; M^H)\right),$$

where M is the given weight matrix.

In this exploration, we will also make use of an equivalent combinatorial characterization of the Bethe free energy as a limit of exact counting problems on graph covers of G. Recall that G' is a covering graph (lift) of G if there exists a homomorphism $h : V_{G'} \to V_G$ such that for any vertex $v \in V_{G'}$, h maps $\partial_{G'}(v)$, the neighborhood of v in G', bijectively onto $\partial_G(h(v))$. A covering graph G' is called a k-cover if $|V_{G'}| = k|V_G|$.

Theorem II.1 (Special Case of Vontobel [12])**.** *For every graph G and every matrix $M \in \mathbb{R}^{m\times m}_{\geq 0}$,*

$$\hom_{\mathrm{B}}(G, M) = \limsup_{k\to\infty} \sqrt[k]{\frac{\sum_{G'\in\mathcal{C}_k(G)} \hom(G', M)}{(k!)^{|E_G|}}},$$

where $\mathcal{C}_k(G)$ is the set of all k-covers of G. Note that $|\mathcal{C}_k(G)| = k!^{|E(G)|}$.

Proof. The detailed proof can be found in Theorem 33 of [12]. While the presentation therein is based on so-called normal factor graphs in which edges correspond to variables and nodes correspond to functions, the same argument can be applied, with minor modification for standard factor graphs. To see this, simply replace every variable node in a standard factor graph with an equality node (variables that are adjacent to a single factor are replaced with a half-edge). There is a bijection between graph covers of the new normal factor graph and graph covers of the standard factor graph such that elements in each bijective pair correspond to exactly the same counting problem. □

B. Log-supermodularity

Definition II.2. A non-negative, real-valued function, $g : \mathbb{R}^n \to \mathbb{R}_+$ is log-supermodular (equivalently, multivariate totally positive of order two) if

$$g(\underline{x})g(\underline{y}) \leq g(\underline{x} \wedge \underline{y})g(\underline{x} \vee \underline{y}),$$

for all $\underline{x}, \underline{y} \in \mathbb{R}^n$, where $\underline{x} \vee \underline{y}$ is the componentwise maximum of the vectors $\underline{x}$ and $\underline{y}$ and $\underline{x} \wedge \underline{y}$ is their componentwise minimum. By Topkis' characterization theorem [19], a strictly positive twice continuously differentiable function $g : \mathbb{R}^n \to \mathbb{R}$ is log-supermodular if and only if $\frac{\partial^2 \log g}{\partial x_i \partial x_j} \geq 0$ for all $i \neq j$.

Log-supermodular functions play an important role in the study of correlation inequalities, e.g., in the FKG inequality. Prior work has generalized the Four Functions Theorem of Ahlswede and Daykin [20] to prove a general relationship between log-supermodular MRFs and their graph covers.

Theorem II.3 (Ruozzi [21]). *If G^k is a k-cover of the graph G given by the covering map h and the potential functions $\psi_{ij:(i,j)\in E_G}$ are all log-supermodular, then*

$$Z(G;\phi,\psi)^k \geq Z(G^k;\hat{\phi},\hat{\psi}),$$

where $\hat{\phi}_i = \phi_{h(i)}$ for each $i \in V_{G^k}$ and $\hat{\psi}_{ij} = \psi_{h(i)h(j)}$ for each $(i,j) \in E_{G^k}$.

III. Graph Covers and Weighted Homomorphism Counting

We begin by showing that the Bethe free energy, $\hom_{\mathrm{B}}(\cdot,\cdot)$, always yields an upper bound on the posited lower bound in Sidorenko's conjecture.

Theorem III.1. *For any bipartite graph $G = (A_G, B_G, E_G)$ and arbitrary graph H with corresponding adjacency matrix M^H,*

$$\hom_{\mathrm{B}}(G, M^H) \geq \hom_{\mathrm{S}}(G, M^H).$$

Proof. We will assume that G is connected with no isolated vertices. If not, a similar argument can be made for each connected component, and the results for each component can be combined using the observation that $\hom_{\mathrm{B}}(G, M^H) = \hom_{\mathrm{B}}(G_1, M^H) \cdot \hom_{\mathrm{B}}(G_2, M^H)$ whenever G is the disjoint union of G_1 and G_2. The Bethe free energy is then

$$\begin{aligned}
&\log \hom_{\mathrm{B}}(G, M^H) = \\
&\sup_{\tau\in\mathcal{T}} \Bigg[\sum_{\substack{i\in A_G, j\in B_G\\(i,j)\in E_G}} \sum_{x_i,x_j} \tau_{ij}(x_i,x_j) \log M^H_{x_i,x_j} \\
&\quad - \sum_{i\in A_G\cup B_G} \sum_{x_i} \tau_i(x_i)\log\tau_i(x_i) \\
&\quad - \sum_{\substack{i\in A_G, j\in B_G\\(i,j)\in E_G}} \sum_{x_i,x_j} \tau_{ij}(x_i,x_j) \log \frac{\tau_{ij}(x_i,x_j)}{\tau_i(x_i)\tau_j(x_j)} \Bigg].
\end{aligned}$$

For all $(i,j) \in E_G$, let $\tau'_{i,j}(x_i,x_j) \triangleq \frac{M^H_{x_i,x_j}}{2|E_H|}$ for all x_i, x_j, and for all $i \in A_G \cup B_G$, let $\tau'_i(x_i) \triangleq \frac{\deg_H(x_i)}{2|E_H|}$ for all x_i. We have that

$$\begin{aligned}
&\log \hom_{\mathrm{B}}(G, M^H) \\
&\overset{(a)}{\geq} \sum_{\substack{i\in A_G, j\in B_G\\(i,j)\in E_G}} \sum_{x_i,x_j} \tau'_{ij}(x_i,x_j) \log M^H_{x_i,x_j} \\
&- \sum_{i\in A_G\cup B_G} \sum_{x_i} \tau'_i(x_i)\log\tau'_i(x_i) \\
&- \sum_{\substack{i\in A_G, j\in B_G\\(i,j)\in E_G}} \sum_{x_i,x_j} \tau'_{ij}(x_i,x_j) \log \frac{\tau'_{ij}(x_i,x_j)}{\tau'_i(x_i)\tau'_j(x_j)} \\
&\overset{(b)}{=} \sum_{\substack{i\in A_G, j\in B_G\\(i,j)\in E_G}} \sum_{x_i,x_j} \tau'_{ij}(x_i,x_j) \log \tau'_{ij}(x_i,x_j)_{x_i,x_j} \\
&\quad + |E_G|\log(2|E_H|) - \sum_{i\in A_G\cup B_G} \sum_{x_i} \tau'_i(x_i)\log\tau'_i(x_i) \\
&\quad - \sum_{\substack{i\in A_G, j\in B_G\\(i,j)\in E_G}} \sum_{x_i,x_j} \tau'_{ij}(x_i,x_j) \log \frac{\tau'_{ij}(x_i,x_j)}{\tau'_i(x_i)\tau'_j(x_j)} \\
&= |E_G|\log(2|E_H|) - \sum_{i\in A_G\cup B_G} \sum_{x_i} \tau'_i(x_i)\log\tau'_i(x_i) \\
&\quad + \sum_{\substack{i\in A_G, j\in B_G\\(i,j)\in E_G}} \sum_{x_i,x_j} \tau'_{ij}(x_i,x_j) \log[\tau'_i(x_i)\tau'_j(x_j)] \\
&\overset{(c)}{=} |E_G|\log(2|E_H|) \\
&\quad + \sum_{i\in A_G\cup B_G} \sum_{x_i} (\deg_G(i)-1)\tau'_i(x_i)\log\tau'_i(x_i) \\
&\overset{(d)}{\geq} |E_G|\log(2|E_H|) \\
&\quad + \sum_{i\in A_G\cup B_G} \left[(\deg_G(i)-1)\log\frac{1}{|V(H)|}\right] \\
&\overset{(e)}{=} (|A_G\cup B_G| - 2|E_G|)\log|V_H| \\
&\quad + |E_G|\log\hom(K_2, M^H),
\end{aligned}$$

where (a) follows from the observation that τ' is in the local marginal polytope (i.e., it represents local probability distributions that satisfy the marginalization conditions), (b) follows from the observation that $M^H_{x_i,x_j} = 2|E_H|\tau'_{ij}(x_i,x_j)$, (c) follows from the fact that for all edges $(i,j) \in E_G$ and all x_i, $\tau'_i(x_i) = \sum_{x_j}\tau'_{ij}(x_i,x_j)$, (d) follows from the observation that the entropy is maximized by the uniform distribution (and consequently that the negative entropy is minimized there), and (e) uses the fact that $2|E_H| = \sum_{x_i,x_j} M^H_{x_i,x_j} = \hom(K_2, M^H)$. From the definition of Sidorenko's lower bound in (4),

$$\begin{aligned}
\hom_{\mathrm{B}}(G, M^H) &\geq |V_H|^{|V_G|-2|E_G|}\hom(K_2, M^H)^{|E_G|} \\
&= \hom_{\mathrm{S}}(G, M^H).
\end{aligned}$$

□

Using the same reasoning as in the proof of Theorem III.1, we can conclude a similar result for general non-negative weight matrices $M \in \mathbb{R}^{m\times m}_{\geq 0}$.

Corollary III.2. *For any bipartite graph G and arbitrary $M \in \mathbb{R}^{m\times m}_{\geq 0}$,*

$$\begin{aligned}
\hom_{\mathrm{B}}(G, M) &\geq \hom(K_2, M)^{|E_G|} m^{|V_G|-2|E_G|} \\
&= \hom_{\mathrm{S}}(G, M).
\end{aligned}$$

Proof. Set $z \triangleq \sum_{x,y\in[m]} M_{x,y}$. For all $(i,j) \in E_G$ and $x_i, x_j \in [m]$, let $\tau'_{i,j}(x_i,x_j) \triangleq M_{x_i,x_j}/z$, for all

$i \in A_G$ and $x_i \in [m]$, let $\tau_i'(x_i) \triangleq (\sum_{x_j \in [m]} M_{x_i,x_j})/z$, for all $j \in B_G$ and $x_j \in [m]$, let $\tau_j'(x_j) \triangleq (\sum_{x_i \in [m]} M_{x_i,x_j})/z$. We can then apply the same argument except that we replace $2|E_H|$ with z in step (b) of the proof. □

Remark III.3. While there exist situations in which equality is achieved in Theorem III.1, e.g., the trivial case in which G is a disjoint union of single edges, $\hom_{\mathrm{B}}$ is often significantly larger than $\hom_{\mathrm{S}}$. For example, let M^{C_4} be the adjacency matrix of the 4-cycle. We have $\hom(K_{3,3}, M^{C_4}) = 164$ while $\hom_{\mathrm{S}}(K_{3,3}, M^{C_4}) = 8$ and $\hom_{\mathrm{B}}(K_{3,3}, M^{C_4}) = 64$.

IV. Disproof of Conjecture I.2

First, note that Conjecture I.2 is easily seen to be false if we remove the bipartite requirement on the graph G. That is, in general, the Bethe approximation need not yield a lower bound on the desired counting problem. In this section, as a counter example to Conjecture I.2, we show that there exists a bipartite graph G and a graph H such that

$$\hom(G, M^H) < \hom_{\mathrm{B}}(G, M^H).$$

We will make use of the following construction. First, let G be the graph obtained by gluing together d disjoint 4-cycles together along a common edge. Now, let H' be an r-regular bipartite graph with vertex set $X \cup Y$ such that $|X| = |Y| = n$. Construct a graph H by adding a vertex a to H' and connecting it to every vertex in X, adding a vertex b and connecting it to every vertex in Y, and then adding a self-loop to both a and b. For appropriate choices of n, r, d, these constructions (see Figure 1) yield a counterexample to Conjecture I.2.

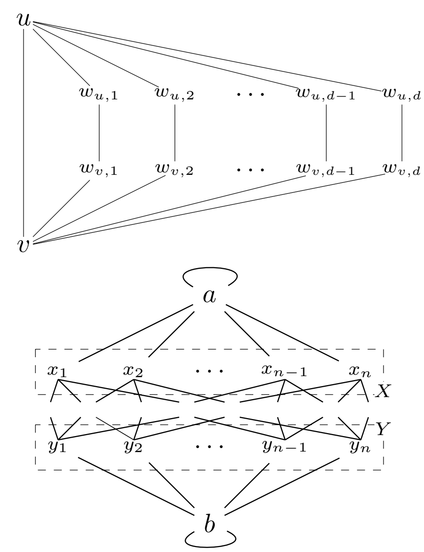

Fig. 1. A bipartite graph G (left) and a target graph H (right) that yield a counter example for Conjecture I.2.

Theorem IV.1. *Let G be the graph obtained by gluing together d disjoint 4-cycles together along a common edge, and let the graph H be obtained from an r-regular bipartite graph $H' = (A, B, E)$ where $|A| = |B| = n$ as described above. Then,*

$$\hom(G, M^H) \leq 4(n+1)^2 \cdot s^d,$$

and

$$\hom_B(G, M^H) \geq \left(\frac{nr}{2}\right)^d,$$

where M^H is the unweighted adjacency matrix of graph H and $s = \max(r(r+1) + n + 1, 2(n+1))$.

Proof. The proof is divided into two parts. First we will show that $\hom(G, M^H) \leq 4(n+1)^2 \cdot s^d$. Let $e = (u, v) \in E_G$ be the common edge of the 4-cyles in G. Let u, w_u, w_v, v be the vertices of a 4-cycle in G (see Figure 1), and let C denote its vertex-induced subgraph. We will bound the number of valid homomorphisms from each 4-cycle into H, and then, by symmetry, use this to bound the total number homomorphisms from G to H.

For a homomorphism $\phi : V_C \to V_H$, we can divide its behavior on the edge (u, v) into three possible cases.

1) If $\phi(u) = \phi(v) = a$, then either $\phi(w_u) = a$, or $\phi(w_v) = a$, or both. So, there are less than $2(n+1)$ such homomorphisms.
2) If $\phi(u) = a$ and $\phi(v) = x_i$ for any $x_i \in X$, then either $\phi(w_v) \in N_H(x_i) \cap Y$ and $\phi(w_u) \in N_H(\phi(w_v)) \cap X$, which yields less than $r(r+1)$ homomorphisms, or $\phi(w_v) = a$ and $\phi(w_u) \in N_H(a)$, which adds $(n+1)$ more. So, all together there are at most $r(r+1) + (n+1)$.
3) If $\phi(u) = x_i$, for any $x_i \in X$ and $\phi(v) = y_j$ for any $y_j \in Y$, then $\phi(w_u) \in N_H(x_i)$ and $\phi(w_v) \in N_H(y_j)$ both yield $(r+1)$ homomorphisms, which implies that there are at most $(r+1)^2$ homomorphisms of this form. Note that $(r+1)^2 \leq r(r+1) + n + 1 \leq s$.

Up to symmetry of H, e.g., by swapping b for a in case 1, these are the only choices for $\phi(u)$ and $\phi(v)$ that can be extended to a valid homomorphism. So, given any ϕ its behavior on the edge (u, v) falls into one of the three cases above, and hence the number of valid homomorphisms for the cycle u, w_1, w_2, v is at most s. Finally, there are at most $2(n+1)$ possible choices for each of $\phi(u)$ and $\phi(v)$ (each of which results in at most s valid homomorphisms) and there are exactly d 4-cycles in the graph G. So, the number of homomorphisms from G to H is at most $4(n+1)^2 \cdot s^d$.

Before proceeding to the second statement of the theorem, consider deleting the edge $e = (u, v)$ from G. If $\phi(u) = a, \phi(v) = b$, $\phi(w_u) \in X$, and $\phi(w_v) \in N_H(\phi(w_u)) \cap Y$, then ϕ is a valid homomorphism. From this we can conclude that $\hom(G - e, H) \geq (nr)^d$. In other words, adding the edge e dramatically decreases the number of homomorphisms from G to H. With this observation, we will proceed to show that $\hom_{\mathrm{B}}(G, M^H) \geq (nr/2)^d$.

In the remaining argument we will analyze the graph covers of G. Let k be an even number. We claim that there are k-covers of G that contradict Conjecture I.2. Denote the d neighbors of u in $V_G \setminus v$ as $w_{u,1}, \dots, w_{u,d}$ and the d neighbors of v in $V_G \setminus u$ as $w_{v,1}, \dots, w_{v,d}$ (see Figure 1, left). Any k-cover of G can be built by the following process. First, from vertices of graph G, take k copies of both u and v and label them $u_1, \dots, u_k$ and $v_1, \dots, v_k$ respectively. Then for all $j \in [k]$ connect u_j and v_j together. Next, for all $i \in [d]$ add k copies of $w_{u,i}$ for each u_j and label them $w^j_{u,i}$, and add k copies of $w_{v,i}$ for each v_j and label them $w^j_{v,i}$. Connect each $w^j_{u,i}$ to u_i and each $w^j_{v,i}$ to v_j. To complete the k-cover, for all $i \in [d]$ and $j \in [k]$, $w^j_{u,i}$ has to be connected to a $w^z_{v,i}$ for some $z \in [d]$ (note that z does not have to be equal to j). Figure 2 visualizes this construction.

We call a cover **good** if each neighbor of u^j with $j \leq k/2$ is paired with a neighbor of v^z with $z > k/2$, and each neighbor of u^j with $j > k/2$ is paired with a neighbor of v^z with $z \leq k/2$. In a good cover, we think of dividing the copies of the vertices in G into two blocks, each of which contains exactly $k/2$ copies of each vertex in G. A cover is good if all edges from the copies of u to the copies of w_u and the edges from the copies of v to the copies of w_v are inside one of the two blocks while the edges from a copy of a w_u to a copy of a w_v must go between the two blocks. See Figure 2 for a visualization of this construction.

By symmetry, there are $(k/2)!^{2d}(k!)^{2d+1}$ good covers. Note that there are $(k!)^{3d+1}$ k-covers of G and that

$$\frac{(k/2)!^{2d}(k!)^{2d+1}}{(k!)^{3d+1}} = \left(\frac{(k/2)!^2}{(k)!}\right)^d > \left(\frac{1}{2^k}\right)^d. \qquad (5)$$

The remainder of the argument is to show that these good covers have *too many* valid homomorphisms into H.

For a good cover G^{k*}, we will lower bound the number of homomorphisms by counting the homomorphisms $\phi : V_{G^{k*}} \to V_H$ such that $\phi(\{u^1, \dots, u^{k/2}\}) = \phi(\{v^1, \dots, v^{k/2}\}) = \{a\}$ and $\phi(\{u^{k/2+1}, \cdots, u^k\}) = \phi(\{v^{k/2+1}, \dots, v^k\}) = \{b\}$. By construction, $\phi(w^i_{u,j}) \notin \{a, b\}$ for any i and j. In H, the vertices a and b have n neighbors in $V_H \setminus \{a, b\}$. So, each of the kd neighbors of the u^is can be mapped to one of n different choices.

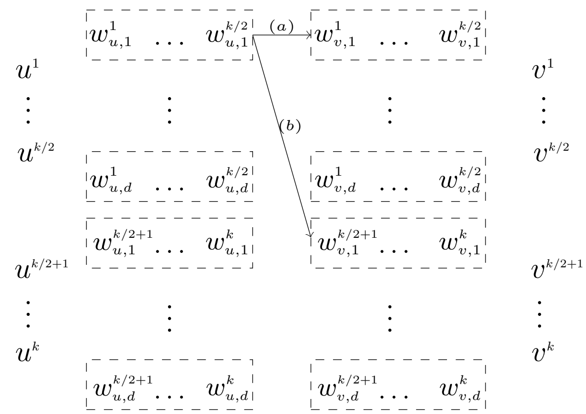

Fig. 2. The vertices of an k-cover of the graph G in the proof of Theorem IV.1. In a valid k-cover, $w^i_{u,j}$ can only pair with $w^\ell_{v,j}$ when $i, \ell \in \{1, \dots, k\}$. In the case of good covers, $w^i_{u,j}$ with $i \leq k/2$ pairs with $w^\ell_{v,j}$ with $\ell > k/2$, i.e., only edges of the form (b) are allowed. As there are $2d$ paired groups, there are $(k/2)!^{2d}$ number of ways to pair every w in a good cover. In an k-cover, there are $k!^d$ number of ways to pair every w.

Consequently, for each of the kd neighbors of v^is there are r feasible choices. As a result, we have $n^{kd}r^{kd}$ of these special homomorphisms. Therefore, the total number of homomorphisms from $V_{G^{k*}}$ to V_H is at least $(nr)^{kd}$. This implies, by Equation (5), that the average homomorphism number among all k-covers is at least $(nr/2)^{kd}$. Hence,

$$\hom_B(G, M^H) \geq \left(\frac{nr}{2}\right)^d.$$

□

Corollary IV.2. *Conjecture I.2 is false.*

Proof. Using Theorem IV.1, any choice of r, n, and d such that

$$\left(\frac{nr}{2}\right)^d > 4(n+1)^2 s^d,$$

yields a counterexample to Conjecture I.2, e.g., $d = 100$, $n = 24$ and $r = 5$. Note, however, that Sidorenko's conjecture holds independent of the choice of n, r, and d: gluing C_4's (or in fact any graphs satisfying Sidorenko's conjecture) along an edge preserves Sidorenko's property [11]. □

V. Special Cases of the Conjecture

While Conjecture I.2 is false in general, in this section, we aim to demonstrate special cases in which Conjecture I.2 is true. These results are still useful: by Theorem III.1 if Conjecture I.2 is true for a specific bipartite graph G, then so is Sidorenko's conjecture. Additionally, the lower bound produced by the Bethe free energy can only improve upon Sidorenko's lower bound in these cases. Key to most of the arguments is a *matrix reformulation* of the Bethe free energy based on the combinatorial characterization.

A. A Matrix Reformulation

In this section, we will show that $\hom_{\mathrm{B}}(G,M) = \limsup_{k\to\infty} \sqrt[k]{\hom(G,M_k)}$ for an appropriately chosen matrix M_k. To begin, notice that every k-cover of a bipartite graph $G=(A_G,B_G,E_G)$, call it G', can be obtained in the following way. First, and by definition of k-covers, $V_{G'}$ consists of k copies of each vertex $i \in V_G$, denoted by $i_1,\dots,i_k$. For each edge $(i,j)\in E_G$, select a permutation $\sigma_{ij}\in\mathcal{S}_k$, where $\mathcal{S}_k$ is the set of all permutations on k elements. Then add the edge $(i_a, j_{\sigma_{ij}(a)})$ to G' for each $a\in[k]$. With this construction, and recalling that $N_{G'}(i)$ denotes the set of neighbors of node i in graph G', we observe that

$$\begin{aligned}\hom(G',M) &= \sum_{x\in[m]^{|V_{G'}|}} \prod_{\substack{i\in A_{G'},j\in B_{G'}\\(i,j)\in E_{G'}}} M_{x_i,x_j}\\ &= \sum_{x\in[m]^{|V_{G'}|}} \prod_{\substack{i\in A_G,j\in B_G\\(i,j)\in E_G}} \prod_{a=1}^{k} M_{x_{i_a},x_{j_{\sigma_{ij}(a)}}}.\end{aligned}$$

The average number of weighted homomorphisms from a k-cover of G to M is then given by the sum over all possible $\sigma_{ij}\in\mathcal{S}_k$ for each $(i,j)\in E_G$, as

$$\begin{aligned}&\sum_{G'\in\mathcal{C}_k(G)} \frac{\hom(G',M)}{(k!)^{|E|}}\\ &= \sum_{G'\in\mathcal{C}_k(G)} \sum_{x\in[m]^{|V_{G'}|}} \prod_{\substack{i\in A_G,j\in B_G\\(i,j)\in E_G}} \frac{1}{k!}\prod_{a=1}^{k} M_{x_{i_a},x_{j_{\sigma_{ij}(a)}}}\\ &= \sum_{x\in[m]^{|V_{G'}|}} \sum_{G'\in\mathcal{C}_k(G)} \prod_{\substack{i\in A_G,j\in B_G\\(i,j)\in E_G}} \frac{1}{k!}\prod_{a=1}^{k} M_{x_{i_a},x_{j_{\sigma_{ij}(a)}}}\\ &\overset{(a)}{=} \sum_{x\in[m]^{|V_{G'}|}} \prod_{\substack{i\in A_G,j\in B_G\\(i,j)\in E_G}} \left[\frac{1}{k!}\sum_{\sigma\in\mathcal{S}_k}\prod_{a=1}^{k} M_{x_{i_a},x_{j_{\sigma(a)}}}\right],\end{aligned}$$

where (a) holds because the choice of k-cover can be done via choosing a permutation for every edge independently. Now, let $\phi:[m]^k\to[m^k]$ be the bijection that sends an element of $[m]^k$ to its position in lexicographical order among all vectors in $[m]^k$. Then for each $I,J\in[m]^k$ we define matrix

$$R_k(M)_{\phi(I),\phi(J)} \triangleq \frac{1}{k!}\sum_{\sigma\in\mathcal{S}_k}\prod_{a=1}^{k} M_{I_a,J_{\sigma(a)}}.$$

The matrix $R_k(M)$ can be equivalently expressed as the product of a matrix depending only on M and a symmetrizing matrix that we will denote as $T^{m,k}$. To see this, let $D\in\mathbb{R}^{m\times m}$ be the identity matrix, and define $D^{\sigma,k}\in\mathbb{R}_{\geq 0}^{m^k\times m^k}$ as

$$D^{\sigma,k}_{\phi(I),\phi(J)} \triangleq \prod_{a=1}^{k} D_{I_a,J_{\sigma(a)}},$$

for all $I,J\in[m]^k$,

Notice that $D^{\sigma,k}\in\mathbb{R}^{m^k\times m^k}$ is a permutation matrix. If $\sigma\in\mathcal{S}_k$ is the identity permutation, then $D^{\sigma,k}=D^{\otimes k}$, the standard k-fold Kronecker product. Define $T^{m,k}\triangleq \frac{1}{k!}\sum_{\sigma\in\mathcal{S}_k} D^{\sigma,k}$. With this definition,

$$\begin{aligned}R_k(M)_{\phi(I),\phi(J)} &= \frac{1}{k!}\sum_{\sigma\in\mathcal{S}_k}\prod_{a=1}^{k} M_{I_a,J_{\sigma(a)}}\\ &= \frac{1}{k!}\sum_{\sigma\in\mathcal{S}_k}\left(M^{\otimes k}D^{\sigma,k}\right)_{\phi(I),\phi(J)}\\ &= \left(M^{\otimes k}T^{m,k}\right)_{\phi(I),\phi(J)}.\end{aligned}$$

The "symmetrizing" matrix $T^{m,k}$ arises naturally in a variety of mathematical applications and has a number of interesting properties that follow from simple algebraic manipulations (see [22] for additional discussion of this operation).

Proposition V.1. *For each $m,k\in\mathbb{N}$,*

- $T^{m,k}\cdot T^{m,k}=T^{m,k}$.
- *For every $M\in\mathbb{R}^{m\times m}$, $T^{m,k}\cdot M^{\otimes k}=M^{\otimes k}\cdot T^{m,k}$.*
- *$T^{m,k}$ is symmetric and doubly stochastic.*
- *For any k vectors $v_1,\dots,v_k\in\mathbb{R}^m$,*

$$T^{m,k}\cdot(v_1\otimes\cdots\otimes v_k) = \frac{1}{k!}\sum_{\sigma\in\mathcal{S}_k} v_{\sigma(1)}\otimes\cdots\otimes v_{\sigma(k)}.$$

As a consequence of the above arguments, we can reduce the counting problem over graph covers to a counting problem over the original graph with a different matrix on each edge. Specfically, we can write

$$\begin{aligned}\sum_{G'\in\mathcal{C}_k(G)} \frac{\hom(G',M)}{(k!)^{|E|}} &= \hom(G,R_k(M))\\ &= \hom(G,M^{\otimes k}T^{m,k}).\end{aligned}\tag{6}$$

Theorem II.1, together with (6), implies the following proposition.

Proposition V.2. *For every bipartite graph G and nonnegative matrix $M\in\mathbb{R}_{\geq 0}^{m\times m}$,*

$$\hom_{\mathrm{B}}(G,M) = \limsup_{k\to\infty}\sqrt[k]{\hom(G,M^{\otimes k}T^{m,k})}.$$

B. A Bilinear Characterization

We can also express the weighted homomorphism counting problem in a bilinear form (see Section VI for an application of this construction). Fix a bipartite graph

G, and observe that $\hom(G, M) = \hom(G, IMI)$ where I is the $m \times m$ identity matrix. We can think of IMI as subdividing each edge of G, in the form of an MRF, into three edges:

$$\begin{aligned}
&\hom(G, IMI)\\
&= \sum_{x\in[m]^{|A_G|}, y\in[m]^{|B_G|}} \prod_{\substack{i\in A_G, j\in B_G\\(i,j)\in E_G}} (IMI)_{x_i,y_j}\\
&= \sum_{\substack{x\in[m]^{|A_G|}, y\in[m]^{|B_G|}\\ r\in[m]^{|E_G|}, s\in[m]^{|E_G|}}} \prod_{\substack{i\in A_G, j\in B_G\\(i,j)\in E_G}} I_{x_i,r_{ij}} M_{r_{ij},s_{ij}} I_{s_{ij},y_j}.
\end{aligned}$$

As was done in Section V-A, let $\phi_k : [m]^k \to [m^k]$ be the bijection that sends an element of $[m]^k$ to its position in lexicographical order among all vectors in $[m]^k$. For ease of notation, we will suppress the dependence on k. For each $x \in [m]^{|A_G|}$ and $r \in [m]^{|E_G|}$, define the matrix

$$V_{\phi(x),\phi(r)} \triangleq \prod_{\substack{i\in A_G, j\in B_G\\(i,j)\in E_G}} I_{x_i,r_{ij}}.$$

Given V, define vector $v \in \mathbb{R}^{m^{|E_G|}}$ such that $v_j = \sum_{i\in[m^{|A_G|}]} V_{i,j}$, for all $j \in [m^{|E_G|}]$. Similarly, for each $s \in [m]^{|E_G|}$ and $y \in [m]^{|B_G|}$ define the matrix

$$W_{\phi(s),\phi(y)} \triangleq \prod_{\substack{i\in B_G, j\in A_G\\(i,j)\in E_G}} I_{s_{ij},y_j}$$

and the vector $w \in \mathbb{R}^{m^{|E_G|}}$ such that $w_i = \sum_{j\in[m^{|B|}]} W_{i,j}$. For a connected bipartite graph $G = (A_G, B_G, E_G)$, $v, w \in \{0,1\}^{m^{|E_G|}}$ with $\sum_i v_i = m^{|A_G|}$ and $\sum_i w_i = m^{|B_G|}$.

Intuitively, the 0-1 vectors v and w act as indicators for valid assignments to the row and column indices of the matrix M for each edge of G. The homomorphism counting problem can then be expressed as the bilinear form

$$\hom(G, M) = v^T M^{\otimes|E_G|} w. \tag{7}$$

Further, suppose that $G = (A, B, E)$ is symmetric in the sense that if $i \in A$ is connected to $j \in B$, then $i \in B$ is connected to $j \in A$. In this case, w and v are equivalent up to a permutation matrix P of the form $D^{\sigma,|E|}$, where D is the $m^{|E|} \times m^{|E|}$ identity matrix and $\sigma \in \mathcal{S}_{|E|}$. So we can express the weighted homomorphism counting problem as

$$\hom(G, M) = v^T M^{\otimes|E_G|} P v. \tag{8}$$

For symmetric bipartite graphs, we can, without loss of generality, assume that v and P are chosen so that P is a symmetric permutation matrix (v and w encode equality constraints on the left and right endpoints of the disjoint union of $|E_G|$ edges, which means that, in the symmetric case, v and w encode the same constraints but in a different ordering of the edges). Note also that the square of any homomorphism number can be put into this form: given a bipartite $G = (A_G, B_G, E_G)$, construct H by forming the disjoint union of G with itself and arranging the partitions of H so that each partition contains one copy of A_G and one copy of B_G that are not connected, i.e., given G_1 and G_2 as two copies of G, define $H = (A_{G_1} \cup B_{G_2}, A_{G_2} \cup B_{G_1}, E_{G_1} \cup E_{G_2})$. This construction gives $\hom(G, M)^2 = \hom(H, M)$.

C. *Quasiconvexity and Weakly Norming Graphs*

Lovász [1] asked under what circumstances the weighted homomorphism counting problem induces a norm. This question gave rise to the study of the so-called weakly norming graphs and the convexity of $\hom(G, \cdot)$.

Definition V.3. Let $\mathcal{W}$ be the space of all two-variable bounded measurable functions on $[0,1]^2$, such that for all $W \in \mathcal{W}$ and $u, v \in [0,1]$, $W(v,u) = W(u,v)$. A graph G is *weakly norming* if for all $W \in \mathcal{W}$,

$$\left(\int_{x\in[0,1]^{V(G)}} \prod_{(i,j)\in E_G} |W(x_i, x_j)| \prod_{i\in V_G} dx_i\right)^{1/|E_G|}$$

is a norm on $\mathcal{W}$ when considered as a function of W, and the integral is with respect to the Lebesgue measure.

All weakly norming graphs are necessarily bipartite and edge transitive [23]. Even cycles, complete bipartite graphs, and hypercubes [7] are known to be weakly norming. However, a complete characterization of weakly norming graphs is still an open problem.

It has been shown that weakly norming graphs satisfy Sidorenko's conjecture, i.e., for a weakly norming graph G, $\hom(G, M^H) \geq \hom_{\mathrm{S}}(G, M^H)$ [7]. Further, it has been shown that a graph G is weakly norming if and only if $\hom(G, \cdot)$ is convex on the set of so-called *signed graphons* [24], [25], a continuous extension of the discrete problem considered herein. An interesting question, then, in our context is whether or not convexity of $\hom(G, \cdot)$, or more generally quasiconvexity, implies anything about its relationship to $\hom_{\mathrm{B}}$. Recall that a function $f : \mathbb{R}^n \to \mathbb{R}$ is quasiconvex if $f(\lambda x + (1-\lambda)y) \leq \max\{f(x), f(y)\}$ for all $x, y \in \mathbb{R}^n$ and all $\lambda \in [0,1]$. The following theorem shows that quasiconvexity of $\hom(G, \cdot)$ combined with permutation invariance on bipartite graphs is enough to conclude that the Bethe free energy yields a lower bound on the weighted homomorphism counting problem.

Theorem V.4. *If G is a bipartite graph such that $\hom(G, M)$ is a quasiconvex function of $M \in \mathbb{R}_{\geq 0}^{m\times m}$,* *then for any matrix $M \in \mathbb{R}^{m\times m}_{\geq 0}$ and any doubly stochastic matrix $S \in \mathbb{R}^{m\times m}_{\geq 0}$,*

$$\hom(G, M) \geq \hom(G, MS).$$

Proof. As G is a bipartite graph, the function $\hom(G, \cdot)$ is permutation invariant, i.e., $\hom(G, MP) = \hom(G, M) = \hom(G, PM)$ for any permutation matrix P. This follows from the observation that permuting the rows or columns of M is equivalent to a change of the variables in the model.

This permutation invariance combined with quasiconvexity implies that $\hom(G, MS) \leq \hom(G, M)$ for any doubly stochastic matrix $S \in \mathbb{R}^{m\times m}$. To see this, recall that by the Birkhoff-von Neumann theorem [26], [27], any doubly stochastic matrix S can be written as $S = \sum_{\sigma\in\mathcal{S}_m} \lambda_\sigma P_\sigma$, where $\sum_{\sigma\in\mathcal{S}_m} \lambda_\sigma = 1$, $\lambda \geq 0$, and P_σ is the permutation matrix corresponding to the permutation σ. We have

$$\begin{aligned}
\hom(G, MS) &= \hom\left(G, M(\sum_{\sigma\in\mathcal{S}_m} \lambda_\sigma P_\sigma)\right) \\
&= \hom\left(G, \sum_{\sigma\in\mathcal{S}_m} \lambda_\sigma M P_\sigma\right) \\
&\overset{(a)}{\leq} \max_{\sigma\in\mathcal{S}_m} \hom(G, MP_\sigma) \\
&\overset{(b)}{=} \hom(G, M)
\end{aligned}$$

where the inequality (a) follows from quasiconvexity of $\hom(G, \cdot)$ and (b) follows from permutation invariance. □

Corollary V.5. *If G is a bipartite graph and $\hom(G, M)$ is a quasiconvex function of $M \in \mathbb{R}^{m\times m}_{\geq 0}$, then $\hom(G, M) \geq \hom_{\mathrm{B}}(G, M)$.*

Proof. Applying Proposition V.2,

$$\begin{aligned}
\hom_{\mathrm{B}}(G, M) &= \limsup_{k\to\infty} \sqrt[k]{\hom(G, M^{\otimes k} T^{m,k})} \\
&\overset{(a)}{\leq} \limsup_{k\to\infty} \sqrt[k]{\hom(G, M^{\otimes k})} \\
&= \limsup_{k\to\infty} \sqrt[k]{\hom(G, M)^k} \\
&= \hom(G, M),
\end{aligned}$$

where the inequality (a) follows from Theorem V.4 and the fact that $T^{m,k}$ is a doubly stochastic matrix. □

VI. Doubly Stochastic Matrices

Corollary V.5 implies that for any bipartite graph G for which $\hom(G, \cdot)$ is quasiconvex and any $m \times m$ doubly stochastic matrix M, $\hom(G, M) \geq \hom(G, (\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T)$, where $\mathbf{1}_m \in \mathbb{R}^{m\times 1}$ has all of its entries equal to 1. In fact, for any $m \times m$ doubly stochastic matrix M, we can check that $\hom_{\mathrm{S}}(G, M) = m^{|V_G|-|E_G|} = \hom(G, (\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T)$. Sidorenko's conjecture for doubly stochastic matrices is then equivalent to the claim that for any bipartite graph G and any $m\times m$ doubly stochastic matrix M,

$$\hom(G, M) \geq \hom(G, (\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T),$$

which is reminiscent of the van der Waerden conjecture for the permanent [28]. In this section, we explore this special case of the conjecture in more detail, i.e., when the domain of $\hom(G, \cdot)$ is the set of $m \times m$ doubly stochastic matrices. We will denote the convex set of all $m \times m$ doubly stochastic matrices by $\mathcal{D}_m$.

We will provide several interesting observations in this special case. We begin with two propositions demonstrating that $\hom(G, \cdot)$ is monotonically nondecreasing along the line segement starting at $(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T$ and ending at any permutation matrix and that $\hom(G, \cdot)$ attains its maximum over $\mathcal{D}_m$ at the permutation matrices.

Proposition VI.1. *For any bipartite graph $G = (A_G, B_G, E_G)$, any $m \times m$ permutation matrix P, and any $\lambda \in [0, 1]$,*

$$\hom(G, (1-\lambda)(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T + \lambda P) \geq \hom(G, (\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T).$$

Proof. Given $Q \subseteq E_G$, define $G^Q = (A_G, B_G, Q)$. Having graph G_Q we have

$$\begin{aligned}
&\hom(G, (1-\lambda)(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T + \lambda P) \\
&= \sum_{x\in[m]^{|V_G|}} \prod_{\substack{i\in A_G, j\in B_G \\ (i,j)\in E_G}} \left[(1-\lambda)(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T + \lambda P\right]_{x_i,x_j} \\
&= \sum_{x\in[m]^{|V_G|}} \prod_{\substack{i\in A_G, j\in B_G \\ (i,j)\in E_G}} \left[\frac{1-\lambda}{m} + \lambda P_{x_i,x_j}\right] \\
&= \sum_{x\in[m]^{|V_G|}} \sum_{Q\subseteq E_G} f_{G,Q}(\lambda, m) \left[\prod_{\substack{i\in A_G, j\in B_G \\ (i,j)\in Q}} \lambda P_{x_i,x_j}\right],
\end{aligned}$$

where

$$\begin{aligned}
f_{G,Q}(\lambda, m) &= \left[\prod_{\substack{i\in A_G, j\in B_G \\ (i,j)\in E_G\setminus Q}} \frac{1-\lambda}{m}\right] \\
&= \left(\frac{1-\lambda}{m}\right)^{|E_{G\setminus Q}|}.
\end{aligned}$$

Since $f_{G,Q}(\lambda, m)$ does not depend on the value of x, swapping the order of summation results in

$$\begin{aligned}&\hom(G,(1-\lambda)(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T+\lambda P)\\&=\sum_{Q\subseteq E_G} f_{G,Q}(\lambda,m)\lambda^{|Q|}\hom(G^Q,P)\\&\overset{(a)}{\geq}\sum_{Q\subseteq E_G} f_{G,Q}(\lambda,m)\lambda^{|Q|}\hom(G^Q,(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T)\\&\overset{(b)}{=}\sum_{Q\subseteq E_G}(1-\lambda)^{|E_G\setminus Q|}\lambda^{|Q|}\hom(G,(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T)\\&=\hom(G,(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T),\end{aligned}$$

where (a) holds because graph G^Q is bipartite and (b) follows from the observation that for any bipartite graph $H = (V_H, E_H)$,

$$\begin{aligned}\hom(H,(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T) &= (\frac{1}{m})^{|E_H|-|V_H|}\\&\leq m^c\\&=\hom(H,P),\end{aligned}$$

where c is the number of connected components of H. Therefore, $\hom(G^Q,(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T)\leq\hom(G^Q,P)$. $\square$

Proposition VI.2. *For any connected bipartite graph $G = (A_G, B_G, E_G)$, any $m\times m$ permutation matrix P, and any $m\times m$ doubly stochastic matrix M,*

$$\hom(G,P)\geq\hom(G,M).$$

Proof. From Equation (7), $\hom(G,M)=v^TM^{\otimes|E_G|}w$ for appropriately chosen v and w. Define the function $g:\mathbb{R}^{m^{|E_G|}\times m^{|E_G|}}\to\mathbb{R}$ as

$$g(A)\triangleq\max_{\sigma_1,\ldots,\sigma_{|E_G|}\in\mathcal{S}_m}\operatorname{trace}(wv^TA\otimes_{k=1}^{|E_G|}P_{\sigma_k}).$$

Note that g is convex and permutation invariant with respect to the subgroup of permutation matrices that can be written as the Kronecker product of $|E_G|$, $m\times m$ permutation matrices. Denote this subgroup as $\mathcal{Q}$. In what follows, let I represent the $m\times m$ identity matrix. We have

$$\begin{aligned}\hom(G,M)&=v^TM^{\otimes|E_G|}w\\&=\operatorname{trace}(wv^TM^{\otimes|E_G|})\\&\overset{(a)}{\leq}g(M^{\otimes|E_G|})\\&\overset{(b)}{\leq}g(I^{\otimes|E_G|})\\&=\max_{\sigma_1,\ldots,\sigma_{|E_G|}\in\mathcal{S}_m}\operatorname{trace}(v^T\otimes_{k=1}^{|E_G|}P_{\sigma_k}w)\\&\overset{(c)}{\leq}m\\&=\hom(G,P),\end{aligned}$$

where (a) follows from the definition of g and the observation that the $m^{|E_G|}\times m^{|E_G|}$ identity matrix is in $\mathcal{Q}$, (b) follows from the fact that $M^{\otimes|E_G|}$ is in the convex hull of the elements of $\mathcal{Q}$ and that g is convex and permutation invariant over $\mathcal{Q}$ (use essentially the same argument as that of Theorem V.4), and (c) follows from the observation that if we put a possibly different $m\times m$ permutation matrix on each edge of G, then there are at most m assignments to the vertices of G that result in a non-zero value. $\square$

Proposition VI.3. *For any bipartite graph $G = (A_G, B_G, E_G)$ that is the disjoint union of trees and any $m\times m$ doubly stochastic matrix M,*

$$\hom(G,M)=\hom(G,(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T).$$

Proof. Without loss of generality, we can assume that G is a connected tree-structured graph. If $|E_G| = 0$, then the result is trivial. Otherwise, it is easy to verify that for any leaf v of G, $\hom(G,M)=\hom(G\setminus v,M)$, where $G\setminus v$ is the graph obtained by deleting the vertex v and its incident edge from G. Repeating this process until a single vertex remains shows that $\hom(G,M)=m$ for any doubly stochastic matrix M. $\square$

Proposition VI.4. *Let C_{2k} be the simple cycle on $2k$ nodes with $k\geq 2$. For all $M\in\mathcal{D}_m$*

$$\hom(C_{2k},M)\geq\hom(C_{2k},(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T),$$

where equality holds if and only if $M=(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T$.

Proof. Since $\hom(C_{2k},M)=\operatorname{trace}((MM^T)^k)$, it is enough to show that $(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T$ is the unique minimizer of $\min_{M\in\mathcal{D}_m}\operatorname{trace}\left((MM^T)^k\right)$. Clearly, $\operatorname{trace}((MM^T)^k)=\sum_{i=1}^m\lambda_i^k$, where $\lambda_1,\lambda_2,\ldots,\lambda_m\geq 0$ are the eigenvalues of the positive semidefinite doubly stochastic matrix MM^T. Since any doubly stochastic matrix has one eigenvalue equal to 1, $\operatorname{trace}((MM^T)^k)=\sum_{i=1}^m\lambda_i^k\geq 1$, where the minimum is attained at the unique rank 1 doubly stochastic matrix, i.e., $MM^T=(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T$. Finally, $M=(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T$ is the unique doubly stochastic solution to $MM^T=(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T$, and hence, the unique minimizer of $\min_{M\in\mathcal{D}_m}\operatorname{trace}\left((MM^T)^k\right)$.

$\square$

Proposition VI.5. *Let G be a bipartite graph. If $\hom(G,\cdot)=\hom(G,(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T)$ has a unique solution on $\mathcal{D}_m$ or G is cycle free, then $\hom(G,M)\geq\hom(G,(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T)$ for all $M\in\mathcal{D}_m$.*

Proof. If G is cycle free the result follows by Proposition VI.3. As a result, we need only consider the case in which $\hom(G,\cdot)=\hom(G,(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T)$ has a unique solution on $\mathcal{D}_m$ and G is a bipartite graph with at least one cycle. Suppose by way of contradiction that there exists some $M \in \mathcal{D}_m$ such that $\hom(G,M) \leq \hom(G,N)$ for all $N \in \mathcal{D}_m$ and that $\hom(G,M) < \hom(G,(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T)$. By Proposition VI.1, we have that for all permutation matrices P,

$$\hom(G,M) \leq \hom(G,(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T) \leq \hom(G,P).$$

By continuity and the assumption that $\hom(G,M) = \hom(G,(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T)$ has a unique solution on $\mathcal{D}_m$, for each permutation matrix P, there exists a unique $t_P \in [0,1]$ such that $M + t_P(P-M) = (\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T$. That is, for every permutation matrix P, the line from M to P must pass through $(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T$. As this cannot be simultaneously true for all permutation matrices unless $M = (\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T$, we encounter a contradiction and the result follows. □

Definition VI.6. For every bipartite graph $G = (A_G, B_G, E_G)$, with $Q \subseteq E_G$, let $\hom_{G,Q}(M,M')$ correspond to the homomorphism counting problem in which the matrix M' is placed on every $e \in Q$ and matrix M is placed on every $e \in E_G \setminus Q$, i.e.,

$$\hom_{G,Q}(M,M') \triangleq \sum_{x\in\{1,\ldots,m\}^{|V|}} \prod_{\substack{i\in A_G, j\in B_G\\(i,j)\in E_G\setminus Q}} M_{x_i,x_j} \prod_{\substack{i\in A_G, j\in B_G\\(i,j)\in Q}} M'_{x_i,x_j}.$$

Now with $U, D \in \mathbb{R}_{\geq 0}^{m\times m}$, the s^{th} directional derivative of $\hom(G,\cdot)$ at U in the direction D, is given by

$$\begin{aligned}
&\frac{d^s}{dt^s}\hom(G,U+tD)\Big|_{t=0}\\
&= \frac{d^s}{dt^s}\left[\sum_{Q\subseteq E_G} t^{|Q|}\hom_{G,Q}(U,D)\right]\Bigg|_{t=0}\\
&= s!\sum_{\substack{Q\subseteq E_G\\|Q|=s}} \hom_{G,Q}(U,D),
\end{aligned}$$

That is, the s^{th} derivative sums over all $\binom{|E_G|}{s}$ edge subsets of size s in which the matrix on the chosen s edges is set to D while the matrix U is placed on the remaining edges.

Proposition VI.7. *For all bipartite G with at least one cycle, $\hom(G,\cdot)$ has a local minimum at $(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T$ over $\mathcal{D}_m$.*

Proof. Fix a doubly stochastic matrix $M \in \mathcal{D}_m$ such that $M \neq (\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T$. We will use the general derivative test to show that $f(t) \triangleq \hom(G,(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T + t[M - (\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T])$ has a local minimum at the point $t=0$. The s^{th} derivative of $f(t)$ is

$$\frac{d^s}{dt^s}f(t) = s!(\frac{1}{m})^{|E_G|-s}\sum_{\substack{Q\subseteq E_G\\|Q|=s}}\sum_{x_{A_G\cup B_G}}\prod_{\substack{i\in A_G, j\in B_G\\(i,j)\in Q}}(M_{x_i,x_j}-\frac{1}{m}). \tag{9}$$

Now let k, an even integer, be the length of the shortest cycle in G. From Proposition VI.3 and (9) the first $k-1$ derivatives of $\hom(G,(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T)$ in the $M - (\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T$ direction are equal to zero. Intuitively, placing matrix $(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T$ on an edge effectively deletes that edge from the graph as $(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T$ is a constant.

For the k^{th} derivative, each of the subsets of k edges corresponds to either a single even cycle or a forest, and by assumption, there must be at least one k-cycle among them. We can then apply the same reasoning as Proposition VI.4 to conclude that the k^{th} derivative is strictly positive. The result then follows by the general derivative test. □

We note that for general, non-bipartite graphs G, the matrix $(\frac{1}{m})\mathbf{1}_m\mathbf{1}_m^T$ need not be a local minimum.

VII. Conclusion

We investigated lower bounds for the weighted homomorphism counting problem on bipartite graphs via the BFE and graph covers. When the BFE yields a lower bound for a given bipartite graph, it can only improve over the lower bound conjectured by Sidorenko. We showed that while this does indeed happen in several special cases, it need not happen for every biparite graph. Indeed, we constructed an explicit counterexample for general bipartite graphs. We believe that the matrix and bilinear characterizations of the $\hom$ and $\hom_{\mathrm{B}}$ may be useful for establishing lower bounds in other scenarios. A similar symmetrizing construction can be made for general MRFs, i.e., those that contain potentials that depend on more than a pair of variables, but we leave the details of such a construction for future work.

Appendix

In this appendix, we provide alternative proofs for the lower bound in the case of single cycles and complete bipartite graphs (that do not rely on the notion of weakly norming). These arguments are based on the notion of majorization.

Definition A.1. A vector $v \in \mathbb{R}^n$ is majorized by a vector $w \in \mathbb{R}^n$, written $v \prec w$, if

$$\sum_{i=1}^{k} v_{[i]} \leq \sum_{i=1}^{k} w_{[i]}, \quad \text{for all } k \in \{1,\ldots,n\} \tag{10}$$

and

$$\sum_{i=1}^{n} v_{[i]} = \sum_{i=1}^{n} w_{[i]},$$

where $w_{[i]}$ denotes the i^{th} largest entry of w.

Equivalently, $v \prec w$ if and only if there exists a doubly stochastic matrix D such that $v = Dw$. The vector v is said to be weakly majorized by w, denoted $v \prec_w w$, if only condition (10) holds. Moreover, if $f : \mathbb{R} \to \mathbb{R}$ is a convex function, then $v \prec w$ implies that $f(v) \prec_w f(w)$, where $f(v)$ denotes the vector obtained by applying f to each component of v [29].

Proposition A.2. *For every $n > 1$ and every $M \in \mathbb{R}_{\geq 0}^{m\times m}$, $\hom(C_{2n}, M) \geq \hom_{\mathrm{B}}(C_{2n}, M)$, where C_{2n} is the simple cycle on $2n$ nodes.*

Proof. Observe that, for a single cycle, $\hom(C_{2n}, M) = \mathrm{tr}((MM^T)^n)$. From Proposition V.2, it suffices to show that $\hom(C_{2n}, M^{\otimes k}T^{m,k}) \leq \hom(C_{2n}, M)^k$ for each $k > 1$. Consider,

$$\begin{aligned}\hom(C_{2n}, M^{\otimes k}T^{m,k}) &\overset{(a)}{=} \mathrm{tr}\left(\left(M^{\otimes k}M^{\otimes k^T}\right)^n T^{m,k}\right) \\ &\overset{(b)}{\leq} \mathrm{tr}\left(\left(\left(MM^T\right)^{\otimes k}\right)^n\right) \\ &\overset{(c)}{=} \mathrm{tr}\left(\left(MM^T\right)^n\right)^k \\ &= \hom(C_{2n}, M)^k,\end{aligned}$$

where (a) follows from Proposition V.1, (b) follows from a standard majorization argument on the eigenvalues of positive semidefinite matrices, see H.1.g. Marshall and Olkin [29], and (c) is a consequence of the observation that $\hom(G, M^{\otimes k}) = \hom(G, M)^k$ for all graphs G. □

Proposition A.3. *For every complete bipartite graph $K_{a,b}$ and every $M \in \mathbb{R}_{\geq 0}^{m\times m}$, $\hom(K_{a,b}, M) \geq \hom_{\mathrm{B}}(K_{a,b}, M)$.*

Proof. The proof considers a reformulation of the counting problem in which the variables in one of the parts of $K_{a,b} = (A, B, E)$ have been summed out. For $N \in \mathbb{R}_{\geq 0}^{m\times m}$,

$$\begin{aligned}\hom(K_{a,b}, N) &= \sum_{x_B\in[m]^b} \sum_{x_A\in[m]^a} \prod_{i\in A} \prod_{j\in B} N_{x_i,x_j} \\ &= \sum_{x_B\in[m]^b} \prod_{i\in A} \left(\sum_{x_i\in[m]} \prod_{j\in B} N_{x_i,x_j} \right) \\ &= \sum_{x_B\in[m]^b} \left(\sum_{y\in[m]} \prod_{j\in B} N_{y,x_j} \right)^a .\end{aligned}$$

We can think of the product $\prod_{j\in B} N_{y,x_j}$ as a vector indexed by assignments to the variables x_B, i.e.,

$$\prod_{j\in B} N_{y,x_j} = \left[(N_{y,:})^{\otimes b}\right]_{x_B},$$

where $N_{y,:}$ denotes the y^{th} row of N. Now, consider $\hom(K_{a,b}, M^{\otimes k}T^{m,k})$ for some $k > 1$. Following the above argument and substituting $M^{\otimes k}T^{m,k}$ for N yields

$$\begin{aligned}&\sum_{y\in[m^k]} \prod_{j\in B} (M^{\otimes k}T^{m,k})_{y,x_j} \\ &\quad= \sum_{y\in[m^k]} \left[\left(M_{y,:}^{\otimes k}T^{m,k}\right)^{\otimes b}\right]_{x_B} \\ &\quad= \sum_{y\in[m^k]} \left[\left(M_{y,:}^{\otimes k}\right)^{\otimes b}\left(T^{m,k}\right)^{\otimes b}\right]_{x_B} \\ &\quad= \left[\sum_{y\in[m^k]} \left(M_{y,:}^{\otimes k}\right)^{\otimes b}\left(T^{m,k}\right)^{\otimes b}\right]_{x_B}\end{aligned}$$

for each $x_B \in [m^k]^b$. Since $T^{m,k}$ is doubly stochastic, so is $\left(T^{m,k}\right)^{\otimes b}$, and we have

$$\sum_{y\in[m^k]} \left(M_{y,:}^{\otimes k}\right)^{\otimes b}\left(T^{m,k}\right)^{\otimes b} \prec \sum_{y\in[m^k]} \left(M_{y,:}^{\otimes k}\right)^{\otimes b}.$$

As $f(x) = x^a$ is a convex function for all $a > 1$, $x \geq 0$, raising each component to the a^{th} power preserves weak majorization.

$$\begin{aligned}&\hom(K_{a,b}, M^{\otimes k}T^{m,k}) \\ &\quad= \sum_{x_B\in[m^k]^b} \left[\sum_{y\in[m^k]} \left(M_{y,:}^{\otimes k}\right)^{\otimes b}\left(T^{m,k}\right)^{\otimes b}\right]_{x_B}^a \\ &\quad\leq \sum_{x_B\in[m^k]^b} \left[\sum_{y\in[m^k]} \left(M_{y,:}^{\otimes k}\right)^{\otimes b}\right]_{x_B}^a \\ &\quad= \hom(K_{a,b}, M^{\otimes k}),\end{aligned}$$

which, in conjunction with Proposition V.2, yields the desired result.

An stronger version of Proposition A.3 for the case in which M is an adjacency matrix follows from observations of Galvin and Tetali [2]. In particular, their results imply that $\hom(G', M^H) \leq \hom(K_{a,b}, M^H)^k$ for any k-cover G' of $K_{a,b}$ and any graph H with adjacency matrix M^H. The above proof argues the result for the k-covers on average but is somewhat simpler than that of [2]. □

References

[1] L. Lovász, *Large networks and graph limits*. American Mathematical Soc., 2012, vol. 60.

[2] D. Galvin and P. Tetali, "On weighted graph homomorphisms," *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, vol. 63, pp. 97–104, 2004.

[3] C. Borgs, J. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi, "Counting graph homomorphisms," in *Topics in discrete mathematics*. Springer, 2006, pp. 315–371.

[4] M. Simonovits, "Extremal graph problems, degenerate extremal problems, and supersaturated graphs," in *Progress in Graph Theory*. NY Academy Press, 1984, pp. 419–437.

[5] A. F. Sidorenko, "Inequalities for functionals generated by bipartite graphs," *(Russian) Diskret. Mat.*, vol. 3, pp. 50–65, 1991.

[6] ——, "A correlation inequality for bipartite graphs," *Graphs and Combinatorics*, vol. 9, pp. 201–204, 1993.

[7] H. Hatami, "Graph norms and Sidorenko's conjecture," *Israel Journal of Mathematics*, vol. 175, no. 1, pp. 125–150, 2010.

[8] D. Conlon, J. Fox, and B. Sudakov, "An approximate version of sidorenko's conjecture," *Geometric and Functional Analysis*, vol. 20, no. 6, pp. 1354–1366, 2010.

[9] J. Li and B. Szegedy, "On the logarithimic calculus and Sidorenko's conjecture," *arXiv preprint arXiv:1107.1153*, 2011.

[10] J. H. Kim, C. Lee, and J. Lee, "Two approaches to sidorenko's conjecture," *Transactions of the American Mathematical Society*, vol. 368, no. 7, pp. 5057–5074, 2016.

[11] B. Szegedy, "An information theoretic approach to Sidorenko's conjecture," *arXiv preprint arXiv:1406.6738*, 2014.

[12] P. O. Vontobel, "Counting in graph covers: A combinatorial characterization of the Bethe entropy function," *Information Theory, IEEE Transactions on*, Jan. 2013.

[13] L. Gurvits, "Unleashing the power of Schrijver's permanental inequality with the help of the Bethe approximation," *ArXiv e-prints*, 2011.

[14] D. Straszak and N. K. Vishnoi, "Belief propagation, Bethe approximation and polynomials," *IEEE Transactions on Information Theory*, 2019.

[15] N. Ruozzi, "Beyond log-supermodularity: Lower bounds and the Bethe partition function," in *Proceedings of the Twenty-Ninth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-13)*. Corvallis, Oregon: AUAI Press, 2013, pp. 546–555.

[16] ——, "The Bethe partition function of log-supermodular graphical models," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2012, pp. 117–125.

[17] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*, ser. Adaptive computation and machine learning. MIT Press, 2009.

[18] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Constructing free-energy approximations and generalized belief propagation algorithms," *IEEE Transactions on information theory*, vol. 51, no. 7, pp. 2282–2312, 2005.

[19] D. M. Topkis, "Minimizing a submodular function on a lattice," *Operations Research*, vol. 26, no. 2, pp. 305–321, 1978.

[20] R. Ahlswede and D. E. Daykin, "An inequality for the weights of two families of sets, their unions and intersections," *Probability Theory and Related Fields*, vol. 43, no. 3, pp. 183–185, 1978.

[21] N. Ruozzi, "A lower bound on the partition function of attractive graphical models in the continuous case," in *Artificial Intelligence and Statistics*, 2017, pp. 1048–1056.

[22] M. Marcus and H. Minc, "Generalized matrix functions," *Transactions of the American Mathematical Society*, vol. 116, pp. 316–329, 1965.

[23] A. Sidorenko, "Weakly norming graphs are edge-transitive," *Combinatorica*, vol. 40, no. 4, pp. 601–604, 2020.

[24] J. Lee and B. Schülke, "Convex graphon parameters and graph norms," *Israel Journal of Mathematics*, vol. 242, no. 2, pp. 549–563, 2021.

[25] M. Doležal, J. Grebík, J. Hladkỳ, I. Rocha, and V. Rozhoň, "Cut distance identifying graphon parameters over weak* limits," *Journal of Combinatorial Theory, Series A*, vol. 189, p. 105615, 2022.

[26] G. Birkhoff, "Three observations on linear algebra," in *Univ. Nac. Tacuman Rev.*, ser. A, vol. 5, 1946, pp. 147–151.

[27] J. von Neumann, "A certain zero-sum two-person game equivalent to an optimal assignment problem," *Ann. Math. Studies*, vol. 28, pp. 5–12, 1953.

[28] G. P. Egorychev, "Solution of van der Waerden's permanent conjecture," *Adv. Math*, vol. 42, pp. 299–305, 1981.

[29] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications*. New York: Academic Press, 1979.

Péter Csikvári recieved his Ph.D. from Alfréd Rényi Institute of Mathematics, Budapest, Hungary, in 2011. He then held a postdoc position at the same university and later another postdoc at MIT, Cambridge, MA. He currently is a researcher of the Alfréd Rényi Institute of Mathematics. He is also an associate professor of the Eötvös Loránd University, Budapest, Hungary.

Nicholas Ruozzi received the B.S. in computer science from Cornell University, Ithaca, NY in 2006 and the Ph.D. in computer science from Yale University, New Haven, CT in 2011. He was a postdoctoral researcher in the Communication Theory Laboratory at Ecole Polytechnique Fédérale de Lausanne (EPFL) in Lausanne, Switzerland from 2011-2013 and then a postdoctoral researcher and adjunct assistant professor at Columbia University from 2013-2015. He is currently an associate professor of computer science at the University of Texas at Dallas. His research interests are in optimization, machine learning, and inference.

Shahab Shams received his B.S. in computer science from Shiraz University in 2013, and Masters in computer science from Sharif University of Technology, Tehran, Iran, in 2016. He then joined University of Texas at Dallas as a Ph.D. student at 2017.