The Homomorphism Domination Exponent

Swastik Kopparty and Benjamin Rossman

June 18, 2009

Abstract

We initiate a study of the *homomorphism domination exponent* of a pair of graphs F and G, defined as the maximum real number c such that $|\mathsf{Hom}(F,T)| \geqslant |\mathsf{Hom}(G,T)|^c$ for every graph T. The problem of determining whether $\mathsf{HDE}(F,G) \geqslant 1$ is known as the *homomorphism domination problem* and its decidability is an important open question arising in the theory of relational databases. We investigate the combinatorial and computational properties of the homomorphism domination exponent, proving upper and lower bounds and isolating classes of graphs F and G for which $\mathsf{HDE}(F,G)$ is computable. In particular, we present a linear program computing $\mathsf{HDE}(F,G)$ in the special case where F is chordal and G is series-parallel.

Contents

1 Introduction

A well known corollary of the Kruskal-Katona theorem states that a graph with e edges can have at most $e^{3/2}$ triangles. More generally one may ask: given two graphs F and G, if we know that a third graph T has a copies of F as a subgraph, what can we say about the number of copies of G in T? This paper is an attempt to pursue a systematic study of a general question of this type.

For (directed) graphs F and G, a *homomorphism from F to G* is a function φ from the vertices of F to the vertices of G such that for any edge (u, v) of F, the pair $(\varphi(u), \varphi(v))$ is an edge of G. The set of all homomorphisms from F to G is denoted $\mathsf{Hom}(F, G)$, its cardinality is denoted $\hom(F, G)$, and we write $F \to G$ if $\hom(F, G) \geqslant 1$.

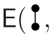
Given a graph T, one can consider the profile of its "subgraph counts" given by the numbers $\hom(F, T)$, as F varies over all finite graphs. The set of all possible profiles encodes much information about the local stucture of graphs. This motivates the following central meta-question in graph theory: find all relations that the numbers $\hom(F_1, T), \ldots, \hom(F_t, T)$ must satisfy in every graph T. Unfortunately, a satisfactory understanding of these relations has thus far been elusive. This failure is explained by the following simple but striking result (due to Ioannidis and Ramakrishnan [IR95], discovered in the context of theoretical databases): given graphs $F_1, \ldots, F_t$ and integers $a_1, \ldots, a_t$, it is undecidable whether for all graphs T, the following inequality holds:

$$\sum_{i=1}^{t} a_i \hom(F_i, T) \geqslant 0.$$

The undecidability already holds if we restrict $t = 13$. Thus, one cannot hope to fully understand the relative magnitudes of subgraph counts of even just 13 graphs at a time! Given this unfortunate fact, we set our sights a little lower, and attempt to study the relative homomorphism numbers from two graphs.

For graphs F and G such that $F \to G$, the *homomorphism domination exponent* of F and G, denoted $\mathsf{HDE}(F, G)$, is defined as the maximal real number c such that $\hom(F, T) \geqslant \hom(G, T)^c$ for all "target" graphs T. The HDE is a parameter encoding deep aspects of the local structure of graphs, and we believe that it is worthy of further study. As a concrete goal, here we consider the question of computing $\mathsf{HDE}(F, G)$ given graphs F and G.

Another motivation for the HDE comes from the theory of databases. The *containment problem for conjunctive queries* (*under multiset semantics*), a problem of much importance in database theory, is equivalent to the *homomorphism domination problem* in graph theory which asks, given graphs F and G, whether $\hom(F, T) \geqslant \hom(G, T)$ for all graphs T. The homomorphism domination exponent is a quantitative version of the homomorphism domination problem (or the conjunctive query containment problem); note that the homomorphism domination problem is simply the question whether $\mathsf{HDE}(F, G) \geqslant 1$.

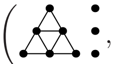
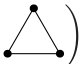
Many classical inequalities involving graphs can be formulated as statements about the homomorphism domination exponent. For example, the Kruskal-Katona theorem implies that in any simple graph G with e edges and t triangles, $e^3 \geqslant t^2$. This may be expressed compactly as $\mathsf{HDE}($, $) \geqslant 2/3$. Similarly, if G has n vertices, e edges and f four-cycles (counted in the sense of homomorphisms), then a result of Kővári, Sós and Turán [KST54] implies that $f \geqslant (e/n)^4$. This can be summarized by the statement $\mathsf{HDE}(C_4 +$ ••••, $) \geqslant 4$. In Section 1.3 we give an overview of known results from extremal combinatorics that imply general bounds on the homomorphisms domination exponent.

Our principal objective in this paper is to give algorithms for computing and bounding the homomorphism domination exponent. We introduce new combinatorial techniques for proving inequalities between homomorphism numbers and establishing their tightness.

1.1 Overview of Results

We prove a lower bound on $\mathsf{HDE}(F, G)$ when F is chordal and G is any graph. This lower bound has the form of a linear program over the convex set of G-polymorphic functions (defined in Section 2.3). In the special case where F is chordal and G is series-parallel, this linear program computes $\mathsf{HDE}(F, G)$ exactly. A relaxation of this linear program turns out to be an upper bound on $\mathsf{HDE}(F, G)$ for all graphs F and G. These results are stated formally in Section 3.

Our bounds yield several new inequalities for graph homomorphism numbers. For instance:

$\mathsf{HDE}\Bigg($ ,  $\Bigg) = \dfrac{5}{2}$,

$$\mathsf{HDE}\big(\text{any directed tree of size } n, \text{ the directed } n\text{-cycle } \vec{C}_n\big) = 1.$$

Let P_n denote the undirected path of size n (with n vertices and $n-1$ edges). Our main theorem implies:

$$\begin{aligned} &\mathsf{HDE}(P_m, P_n) = 1 && \text{when } m \geqslant n, \\ &\mathsf{HDE}(P_m, P_n) = m/n && \text{when } m \leqslant n \text{ and } m \text{ is odd.} \end{aligned}$$

However, when $m \leqslant n$ and m is even, the value of $\mathsf{HDE}(P_m, P_n)$ is slightly less than m/n (by an amount that depends on $n \mod m$):

$$\mathsf{HDE}(P_2, P_n) = 1/\lceil n/2 \rceil,$$

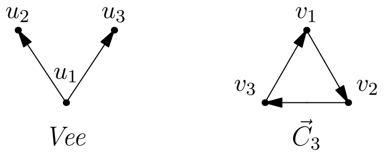
$$\mathsf{HDE}(P_4, P_{4n+i}) = \begin{cases} 1/n & \text{if } i = 0, \\ 2/(2n+1) & \text{if } i = 1, \\ (4n+1)/(4n^2+3n+1) & \text{if } i = 2, \\ 1/(n+1) & \text{if } i = 3. \end{cases}$$

The equation $\mathsf{HDE}(P_4, P_{4n+2}) = (4n+1)/(4n^2+3n+1)$, for instance, was discovered by evaluating the linear program in our main theorem for small values of n and then generalizing (this equation is proved in Theorem 3.4). We remark that finding a closed expression for $\mathsf{HDE}(P_m, P_n)$ for all m and n is an open problem.

1.2 The Method via an Example

We prove our bounds using an approach based on entropy and linear programming. We now briefly illustrate our methods in action on a simple example. The argument is inspired by the entropy proof of Shearer's lemma, often attributed to Radhakrishnan, and its generalizations due to Friedgut and Kahn [FK98, Fri04].

Consider the graphs *Vee* and $\vec{C}_3$ pictured below.



We will prove that $\mathsf{HDE}(\mathit{Vee}, \vec{C}_3) = 1$. (This problem was posed by Erik Vee [Vee06]; different solution and generalization were given by Rossman and Vee [RV06].) As $\hom(\mathit{Vee}, \vec{C}_3) = 3$ and $\hom(\vec{C}_3, \vec{C}_3) = 3$, we have $\mathsf{HDE}(\mathit{Vee}, \vec{C}_3) \leqslant 1$. It remains to show that for all graphs T, $\hom(\mathit{Vee}, T) \geqslant \hom(\vec{C}_3, T)$. To that end, fix an arbitrary graph T such that $\vec{C}_3 \to T$. Pick χ uniformly at random from $\mathsf{Hom}(\vec{C}_3, T)$. For $i = 1, 2, 3$, let $a_i = \chi(v_i)$. Observe that the joint distribution (a_1, a_2, a_3) is uniform on a subset of $T \times T \times T$ of size $\hom(\vec{C}_3, T)$. Thus $\mathbb{H}(a_1, a_2, a_3) = \log \hom(\vec{C}_3, T)$. We now prove that $\mathbb{H}(a_1, a_2, a_3) \leqslant \log \hom(\mathit{Vee}, T)$.

By the chain rule of entropy,

$$\mathbb{H}(a_1, a_2, a_3) = \mathbb{H}(a_1) + \mathbb{H}(a_2|a_1) + \mathbb{H}(a_3|a_1, a_2).$$

As conditioning on fewer variables can only increase entropy, we get

$$\mathbb{H}(a_1, a_2, a_3) \leqslant \mathbb{H}(a_1) + \mathbb{H}(a_2|a_1) + \mathbb{H}(a_3|a_2).$$

Now, by cyclic symmetry of a_1, a_2, a_3, we have $\mathbb{H}(a_3|a_2) = \mathbb{H}(a_2|a_1)$. Thus,

$$\mathbb{H}(a_1, a_2, a_3) \leqslant \mathbb{H}(a_1) + 2\mathbb{H}(a_2|a_1). \tag{1}$$

We will now interpret this expression. Consider the distribution (x, y, y') on $T \times T \times T$ defined as follows. First, $x \in T$ is picked according to the distribution of a_1. Next, two independent copies $y, y' \in T$ of a_2 conditioned on $a_1 = x$ are picked. The entropy of (x, y, y') is easily computed:

$$\mathbb{H}(x, y, y') = \mathbb{H}(x) + \mathbb{H}(y|x) + \mathbb{H}(y'|x) = \mathbb{H}(a_1) + \mathbb{H}(a_2|a_1) + \mathbb{H}(a_2|a_1).$$

Thus, we have $\mathbb{H}(a_1, a_2, a_3) \leqslant \mathbb{H}(x, y, y')$ by (1).

Distribution (x, y, y') was constructed so that there is always an edge from x to y as also from x to y'. Thus, every point of $T \times T \times T$ in the support of the distribution of (x, y, y') specifies a unique homomorphism in $\mathsf{Hom}(\mathit{Vee}, T)$, namely the map $u_1 \mapsto x$, $u_2 \mapsto y$ and $u_3 \mapsto y'$. This implies that $\log \hom(\vec{C}_3, T) = \mathbb{H}(a_1, a_2, a_3) \leqslant \log \hom(\mathit{Vee}, T)$, completing the proof.

The proof of our lower bound on $\mathsf{HDE}(F, G)$ for chordal graphs F and arbitrary graphs G follows the same strategy as the argument above. When we want to prove that for all T, $\hom(F, T) \geqslant \hom(G, T)^c$, we start with a uniform distribution on $\mathsf{Hom}(G, T)$. We analyze its entropy and compare it with the entropy of several auxiliary distributions that we construct on $\mathsf{Hom}(F, T)$. The construction of the auxiliary distributions, as well as the analysis and comparisons of entropies are guided by a linear program.

1.3 Related Work

Several computational problems closely related to the computability of the homomorphism domination exponent are known to be undecidable. Validity of linear inequalities involving homomorphism numbers was shown to be undecidable by [IR95] via a reduction from Hilbert's 10th problem on solvability of integer diophantine equations. The homomorphism domination problem with "inequality constraints" is also known to be undecidable [JKV06].

Inequalities between homomorphism numbers have been extensively studied in extremal combinatorics. For a survey, see [BCL$^+$06a]. Very few general results are known about the homomorphism domination exponent (defined here for the first time, but implicitly studied before). Alon [Alo81] showed that if e is an undirected edge and G is any simple graph, then $\mathsf{HDE}(e, G) = \frac{1}{\rho(G)}$, where $\rho(G)$ is the *fractional edge covering number* of G. This result was reproved and generalized to hypergraphs by Friedgut and Kahn [FK98]. Their argument used Shearer's lemma, which is closely related to the entropy techniques that we use. A wonderful exposition on using entropy and Shearer's lemma to prove classical inequalities can be found in [Fri04]. Galvin and Tetali [GT04], generalizing an argument of Kahn [Kah01], also using entropy techniques, showed that for any n-regular, N-vertex bipartite graph G, $\mathsf{HDE}(K_{n,n}, G) = \frac{2n}{N}$. Finally, a very general approach to inequalities between homomorphism numbers in dense graphs was developed in [BCL$^+$06a, Raz07]. However, it is not known whether this approach leads to algorithms for deciding validity of special families of inequalities between homomorphism numbers.

The entropy arguments that we use differ from the above applications in that we utilize finer information about conditional entropy. The key technical device that enables us to use this information is the construction of auxiliary distributions using conditionally independent copies of the same random variable. This is exemplified in the example of the previous subsection by our definition of the distribution (x, y, y').

Paper Organization. Section 2 introduces the necessary definitions and tools related to graphs and homomorphisms. Our results are formally stated in Section 3. Definitions and auxiliary lemmas on Markov random fields are given in Section 4. Proofs of our main theorems are presented in Sections 5, 6, 7 and 8. We state our conclusions in Section 9.

2 Preliminaries

We first fix some basic notation. For a natural number n, let $[n]$ denote the set $\{1, \ldots, n\}$. The powerset of a set X is denoted by $\wp(X)$. If $\mathcal{S}$ is a family of sets, let $\bigcap \mathcal{S}$ denote the intersection $\bigcap_{S \in \mathcal{S}} S$. We adopt the convention that $\bigcap \emptyset = \emptyset$.

2.1 Graphs and Homomorphisms

Graphs will be finite and directed. Formally, a graph is a pair $G = (V_G, E_G)$ where V_G is a nonempty finite set and E_G is a subset of $V_G \times V_G$. For a subset $A \subseteq V_G$, we denote by $G|_A$ the induced subgraph of G with vertex set A. We denote by $k{\cdot}G$ the disjoint union of k copies of G. The *(categorical) product* $F \times G$ of graphs F and G has vertex set $V_{F \times G} = V_F \times V_G$ and edge set $E_{F\times G} = \{((a, v), (b, w)) : (a, b) \in E_F$ and $(v, w) \in E_G\}$.

A graph G is *simple* if the relation E_G is antireflexive and symmetric, i.e., if $(v, w) \in E_G$ then $v \neq w$ and $(w, v) \in E_G$. Every graph G is associated with a simple graph $\overline{G}$ defined by $V_{\overline{G}} = V_G$ and $E_{\overline{G}} = \{(v, w) : v \neq w$ and $(v, w) \in E_G$ or $(w, v) \in E_G\}$. Whenever we speak of cliques, connectivity, etc., of G, we mean cliques, connectivity, etc., of the associated simple graph $\overline{G}$. In particular, a *clique* in a graph G is a set of vertices $A \subseteq V_G$ such that $(v, w) \in E_G$ or $(w, v) \in E_G$ for all distinct $v, w \in A$. We denote by $\mathsf{Cliques}(G)$ the set of cliques in G and by $\mathsf{MaxCliques}(G)$ the set of maximal cliques in G. The number of connected components of G is denoted by $\mathsf{CC}(G)$.

A *homomorphism* from a graph F to a graph G is a function $\varphi : V_F \longrightarrow V_G$ such that $(\varphi(a), \varphi(b)) \in E_G$ for all $(a, b) \in E_F$. Let $\mathsf{Hom}(F, G)$ denote the set of homomorphisms from F to G and let $\hom(F, G) = |\mathsf{Hom}(F, G)|$. Notation $F \to G$ expresses $\hom(F, G) \geqslant 1$. Under disjoint unions $(+)$ and categorical graph product $(\times)$, $\hom(\_, \_)$ obeys identities

$$\hom(F_1 + F_2, G) = \hom(F_1, G) \cdot \hom(F_2, G),$$
$$\hom(F, G_1 \times G_2) = \hom(F, G_1) \cdot \hom(F, G_2).$$

A graph F is *chordal* if the simple graph $\overline{F}$ contains no induced cycle of size $\geqslant 4$. Chordal graphs are alternatively characterized by the existence of an elimination ordering. A vertex v is *eliminable* in a graph F if the neighborhood of v is a clique in F. An enumeration $v_1, \dots, v_n$ of V_F is an *elimination ordering* for F if v_j is eliminable in $F|_{\{v_1,\dots,v_j\}}$ for all $j \in [n]$. By a well-known characterization, a graph F is chordal if and only if it has an elimination ordering.

A 2-*tree* is a chordal graph with clique number at most 3 (i.e., containing no K_4). A graph G is *series-parallel* if G is a subgraph of some 2-tree.

2.2 The Homomorphism Domination Exponent

We now formally define the homomorphism domination exponent.

Definition 2.1 (Homomorphism Domination Exponent)**.** For graphs F and G such that $F \to G$,[1] the *homomorphism domination exponent* $\mathsf{HDE}(F, G)$ is defined by

$$\mathsf{HDE}(F, G) = \sup\big\{c \in \mathbb{R} : \hom(F, T) \geqslant \hom(G, T)^c \text{ for all graphs } T\big\}.$$

We write $F \succcurlyeq G$ and say F *homomorphism-dominates* G if $\mathsf{HDE}(F, G) \geqslant 1$.

The following dual expression for $\mathsf{HDE}(F, G)$ is often useful:

$$\mathsf{HDE}(F, G) = \inf_{T\,:\,\hom(G,T)\geqslant 2} \frac{\log \hom(F, T)}{\log \hom(G, T)}. \tag{2}$$

We remark that this inf is not always a min.

The following lemma (proof omitted) lists some basic properties of the homomorphism domination exponent.

---

[1] We do not define $\mathsf{HDE}(F, G)$ whenever $F \not\to G$. However, it might be a reasonable convention to let $\mathsf{HDE}(F, G) = -\infty$.

Lemma 2.2 (Basic Properties of HDE)**.**

(a) *If* $c = \mathsf{HDE}(F,G)$*, then* $\hom(F,T) \geqslant \hom(G,T)^c$ *for all graphs* T*. (That is, we can replace* sup *by* max *in Definition 2.1.)*

(b) *The homomorphism-domination relation* $\succcurlyeq$ *is a partial order on graphs.*

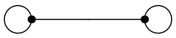
(c) $\mathsf{HDE}(F,H) \geqslant \mathsf{HDE}(F,G) \cdot \mathsf{HDE}(G,H)$*.*

(d) $\mathsf{HDE}(m{\cdot}F, n{\cdot}G) = \frac{m}{n} \cdot \mathsf{HDE}(F,G)$ *for all positive integers* m, n*.*

(e) *If there exists a surjective homomorphism from* F *onto* G*, then* $F \succcurlyeq G$*.*

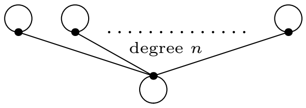
(f) $\mathsf{HDE}(F,G) > 0$ *if and only if* $\bigcup_{\varphi \in \mathsf{Hom}(F,G)} \mathrm{Range}(\varphi) = V_G$*.*

By (2), every graph T with $\hom(G,T) \geqslant 2$ provides an upper bound on $\mathsf{HDE}(F,G)$. By taking specific graphs T_1, T_2 and $(T_{3,n})_{n \geqslant 1}$ in the figure below, we get the following general upper bounds on $\mathsf{HDE}(F,G)$.



T_1

T_2



$T_{3,n}$

Taking $T = T_1$, we get the upper bound $\mathsf{HDE}(F,G) \leqslant |V_F|/|V_G|$. Taking $T = T_2$, we have that $\mathsf{HDE}(F,G) \leqslant \mathsf{CC}(F)/\mathsf{CC}(G)$. A slightly more complicated upper bound follows by taking $T = T_{3,n}$ and letting $n \to \infty$; the result is that $\mathsf{HDE}(F,G)$ is at most the ratio $\alpha(F)/\alpha(G)$ of the independence numbers of F and G, since $\hom(H, T_{3,n})$ grows like $\Theta(n^{\alpha(H)})$ for every graph H.

2.3 G-Polymatroidal Functions

Definition 2.3. For a graph G, let $\mathcal{P}(G)$ and $\mathcal{Q}(G)$ be the following sets of functions from $\wp(V_G)$ to $[0,1]$.

- A function $p : \wp(V_G) \longrightarrow \mathbb{R}$ is *G-polymatroidal* if it satisfies the following four conditions:

(0 at $\emptyset$)	$p(\emptyset) = 0$,
(monotone)	$p(A) \leqslant p(B)$ for all $A \subseteq B \subseteq V_G$,
(submodular)	$p(A \cap B) + p(A \cup B) \leqslant p(A) + p(B)$ for all $A, B \subseteq V_G$,
(G-independent)	$p(A \cap B) + p(A \cup B) = p(A) + p(B)$ for all $A, B \subseteq V_G$ such that $A \cap B$ separates $A \setminus B$ and $B \setminus A$ in G.

  A G-polymatroidal function p is *normalized* if in addition it satisfies:

  (normalized) $p(V_G) = 1$.

- $\mathcal{P}(G)$ denotes the set of normalized G-polymatroidal functions.

- $\mathcal{Q}(G)$ denotes the set of functions $q : \wp(V_G) \longrightarrow \mathbb{R}$ which satisfy:

$$q(\emptyset) = 0, \qquad q(A) \geqslant 0 \text{ for all } A \subseteq V_G, \qquad \sum_{A \subseteq V_G} q(A) \cdot \mathsf{CC}(G|_A) = 1.$$

Example 2.4. Let a, b, c be the vertices of K_3. Then $\mathcal{P}(K_3)$ is the set of convex combinations of eight functions from $\wp(\{a,b,c\})$ to $[0,1]$, which we label as $f_a, f_b, f_{ab}, f_{ac}, f_{bc}, f_{abc}$ (corresponding to the seven nonempty subsets of $\{a,b,c\}$) and f_{RS} ("RS" stands for Ruzsa-Szemeredi, for reasons that will be explained later on), given by the following table:

	$\emptyset$	$\{a\}$	$\{b\}$	$\{c\}$	$\{a,b\}$	$\{a,c\}$	$\{b,c\}$	$\{a,b,c\}$
f_a	0	1	0	0	1	1	0	1
f_b	0	0	1	0	1	0	1	1
f_c	0	0	0	1	0	1	1	1
f_{ab}	0	1	1	0	1	1	1	1
f_{ac}	0	1	0	1	1	1	1	1
f_{bc}	0	0	1	1	1	1	1	1
f_{acb}	0	1	1	1	1	1	1	1
f_{RS}	0	1/2	1/2	1/2	1	1	1	1

We will use the following identity for G-polymatroidal functions when G is chordal.

Lemma 2.5 (Identity for Chordal-Polymatroidal Functions)**.** *If G is chordal, then for every G-polymatroidal function $p : \wp(V_G) \longrightarrow \mathbb{R}$ and every elimination ordering $v_1, \ldots, v_n$ for G,*

$$\begin{aligned} p(V_G) &= \sum_{S \subseteq \mathsf{MaxCliques}(G)} -(-1)^{|S|} p(\textstyle\bigcap S) \\ &= \sum_{i=1}^{n} p\big(\{\text{neighbors of } v_i \text{ among } v_1, \ldots, v_{i-1}\} \cup \{v_i\}\big) - p\big(\{\text{neighbors of } v_i \text{ among } v_1, \ldots, v_{i-1}\}\big). \end{aligned}$$

Lemma 2.5 is established by a straightforward inductive argument (proof omitted).

3 Results

Our first theorem gives a lower bound on $\mathsf{HDE}(F,G)$ when F is chordal.

Theorem 3.1. *If F is chordal and G is any graph, then*

$$\mathsf{HDE}(F,G) \geqslant \min_{p \in \mathcal{P}(G)} \max_{\varphi \in \mathsf{Hom}(F,G)} \sum_{S \subseteq \mathsf{MaxCliques}(F)} -(-1)^{|S|} \cdot p(\varphi(\textstyle\bigcap S)).$$

Theorem 3.1 is proved by a generalization of the entropy technique illustrated by the example in §1.2.

Our second theorem gives an upper bound on $\mathsf{HDE}(F,G)$ for general graphs F and G.

Theorem 3.2. *For all graphs F and G,*

$$\mathsf{HDE}(F,G) \leqslant \min_{q \in \mathcal{Q}(G)} \max_{\varphi \in \mathsf{Hom}(F,G)} \sum_{A \subseteq V_G} q(A) \cdot \mathsf{CC}(F|_{\varphi^{-1}(A)})$$

The next theorem establishes that Theorem 3.1 is tight in the special case where G is series-parallel.

Theorem 3.3. *If F is chordal and G is series-parallel, then*

$$\mathsf{HDE}(F,G) = \min_{p \in \mathcal{P}(G)} \max_{\varphi \in \mathsf{Hom}(F,G)} \sum_{S \subseteq \mathsf{MaxCliques}(F)} -(-1)^{|S|} \cdot p(\varphi(\textstyle\bigcap S)).$$

The final theorem (mentioned in the introduction) is an example of an interesting HDE computation discovered using the linear program of Theorem 3.3.

Theorem 3.4. $\mathsf{HDE}(P_4, P_{4n+2}) = \dfrac{4n+1}{4n^2+3n+1}$

Theorems 3.1, 3.2, 3.3 and 3.4 are respectively proved in Sections 5, 6, 7 and 8.

Discussion 1. Tightness of our lower and upper bounds

The HDE upper bound of Theorem 3.2 is not tight for all pairs of graphs. For instance, $F = C_4 + 2{\cdot}K_1$ (an undirected 4-cycle plus two isolated vertices) and $G = K_2$, it holds that $\mathsf{HDE}(F, G) = 8/3$, while Theorem 3.2 only implies $\mathsf{HDE}(F, G) \leqslant 3$. However, we can show that Theorem 3.2 is tight when (the underlying simple graphs of) F and G are forests.

We do not have any example of a chordal graph F and a graph G for which the HDE lower bound of Theorem 3.1 is not tight. However, there are reasons to believe that the tightness of this lower bound is not the question. Recall that the linear program in Theorem 3.1 has domain $\mathcal{P}(G)$, the set of normalized G-polymatroidal functions. In fact (as will obvious from the proof of Theorem 3.1), we can replace $\mathcal{P}(G)$ with the subset $\{h_X : X \in \mathsf{MRF}(G)\}$ of normalized entropic functions of Markov random fields over G (defined in the next section). Let $\mathcal{E}(G)$ denote the closure of $\{h_X : X \in \mathsf{MRF}(G)\}$ in $\mathbb{R}^{V_G}$. The set $\mathcal{E}(G)$, whose member are called *G-entropic functions*, is a convex subset of $\mathcal{P}(G)$ and a well-studied object in information theory. When $|V_G| \leqslant 3$, we have $\mathcal{E}(G) = \mathcal{P}(G)$. However, these sets do no coincide in general. For instance, $\mathcal{E}(K_4)$ is a proper subset of $\mathcal{P}(K_4)$ (due the existence of "non-Shannon information inequalities" on 4 random variables); in fact, $\mathcal{E}(\mathcal{K}_4)$ fails even to be a polytope. While it seems unnatural to conjecture that the HDE lower bound of Theorem 3.1 is tight as stated, the same conjecture for the corresponding linear program over $\mathcal{E}(G)$ would appear more reasonable.

Discussion 2. Theorem 3.2 is a linear program relaxation of Theorem 3.1

It is worth pointing that linear program in the HDE upper bound of Theorem 3.2 is (after a linear change of variables) a direct relation of linear program in the HDE lower bound of Theorem 3.1. To see this, consider the invertible linear transformation $L : \mathbb{R}^{\wp(V_G)} \longrightarrow \mathbb{R}^{\wp(V_G)}$, which takes a function $f : \wp(V_G) \longrightarrow \mathbb{R}$ to a function $Lf : \wp(V_G) \longrightarrow \mathbb{R}$ defined by

$$(Lf)(A) = \sum_{B\,:\,A\cup B = V_G} -(-1)^{|A\cap B|} f(B).$$

We need a combinatorial lemma on chordal graphs.

Lemma 3.5. *Suppose F is chordal.*

(a) *For all $A \subseteq V_F$,*

$$\sum_{S \subseteq \mathsf{MaxCliques}(F)} (-1)^{|S|} = \sum_{B\,:\,A\cup B = V_F} (-1)^{|A\cap B|} \mathsf{CC}(F|_B).$$

(b) *For every function $f : \wp(V_F) \longrightarrow \mathbb{R}$,*

$$\sum_{S \subseteq \mathsf{MaxCliques}(F)} -(-1)^{|S|} f(\textstyle\bigcap S) = \sum_{A \subseteq V_F} (Lf)(A) \cdot \mathsf{CC}(F|_A).$$

(c) *For every homomorphism $\varphi : F \longrightarrow G$ and function $g : \wp(V_G) \longrightarrow \mathbb{R}$,*

$$\sum_{S \subseteq \mathsf{MaxCliques}(F)} -(-1)^{|S|} g(\varphi(\bigcap S)) = \sum_{A \subseteq V_G} (Lg)(A) \cdot \mathsf{CC}(F|_{\varphi^{-1}(A)}).$$

Lemma 3.5 can be proved by an inductive argument, or alternatively, using elementary algebraic topology (Euler characteristics of flag complexes associated with chordal graphs). Statement (a) is the essential identity; statement (b) follows directed from (a); statement (c), which is the result we need, is a slight extension of (b).

As an immediate corollary of Lemma 3.5(c), we get:

Corollary 3.6 (Alternative Statement of Theorem 3.1)**.** *If F is chordal and G is any graph, then*

$$\mathsf{HDE}(F,G) \geqslant \min_{q \in L(\mathcal{P}(G))} \max_{\varphi \in \mathsf{Hom}(F,G)} \sum_{A \subseteq V_G} q(A) \cdot \mathsf{CC}(F|_{\varphi^{-1}(A)}).$$

To see that the linear program of Theorem 3.2 is a direct relaxation of the linear program of Theorem 3.1, it suffices to show that $\mathcal{Q}(G) \subseteq L(\mathcal{P}(G))$ for all graphs G, which can be checked by applying L^{-1} to an arbitrary function in $\mathcal{Q}$ and seeing that the resulting function is normalized G-polymatroidal. Indeed, for any $q \in \mathcal{Q}(G)$, the function $L^{-1}q$ is given by $(L^{-1}q)(A) = \sum_{B \subseteq V_G} q(B) \cdot \mathsf{CC}(G|_{\varphi^{-1}(A \cap B)})$, which one can show is normalized G-polymatroidal.

4 Chordal Pullbacks of Markov Random Fields

A *(probability) distribution* over a nonempty finite set Ω is a function $X : \Omega \longrightarrow [0,1]$ such that $\sum_{\omega \in \Omega} X(\omega) = 1$. We denote by $\mathsf{Dist}(X)$ the set of all distributions over Ω. The *support* of X is the set $\mathsf{Supp}(X) = \{\omega \in \Omega : X(\omega) > 0\}$. The *entropy* of X is defined by $\mathbb{H}(X) = \sum_{\omega \in \Omega} -X(\omega) \log X(\omega)$. Since the uniform distribution maximizes entropy among all distributions with a given support, it holds that $\mathbb{H}(X) \leqslant \log |\mathsf{Supp}(X)|$.

For a finite set I, we refer to distributions $X \in \mathsf{Dist}(\Omega^I)$ as called *I-indexed joint distribution (with values in Ω).* We view the coordinates X_i $(i \in I)$ as random variables taking values in Ω. We speak of *independence* and *conditional independence* among random variables X_i. For all $J \subseteq I$, we denote by X_J the *marginal J-indexed joint distribution* $\langle X_j : j \in J \rangle$ viewed as a distribution in $\mathsf{Dist}(\Omega^J)$.

For an I-indexed joint distribution X, we denote by $h_X : \wp(I) \longrightarrow [0,1]$ the *normalized entropy function of X* defined by $h_X(J) = \mathbb{H}(X_J)/\mathbb{H}(X)$. By Shannon's classical information inequalities (see [Yeu06]), the function h_X is monotone and submodular.

For a graph G, a V_G-indexed joint distribution $X \in \mathsf{Dist}(\Omega^{V_G})$ is a *Markov random field over G* if $\mathbb{H}(X_A) + \mathbb{H}(X_B) = \mathbb{H}(X_{A \cup B}) + \mathbb{H}(X_{A \cap B})$ for all $A, B \subseteq V_G$ such that $A \cap B$ separates $A \setminus B$ and $B \setminus A$ in G. By Shannon's information inequalities, for $X \in \mathsf{MRF}(G)$, the function $A \longmapsto \mathbb{H}(X_A)$ is G-polymatroidal (recall Definition 2.3). Hence, assuming $\mathbb{H}(X) > 0$, the normalized entropy function h_X belongs to $\mathcal{P}(G)$. By Lemma 2.5, it follows that

$$\mathbb{H}(X) = \sum_{S \subseteq \mathsf{MaxCliques}(G)} -(-1)^{|S|} \mathbb{H}(X_{\cap S}). \tag{3}$$

We denote by $\mathsf{MRF}(G, \Omega)$ the set of all Markov random fields over G with values in Ω. We write $\mathsf{MRF}(G)$ for the class of all Markov random fields over G. Note that $\mathsf{MRF}(G)$ depends only on the underlying simple graph of G. If G_1 and G_2 are simple graphs such that $V_{G_1} = V_{G_2}$ and $E_{G_1} \supseteq E_{G_2}$, then $\mathsf{MRF}(G_1) \subseteq \mathsf{MRF}(G_2)$, i.e., every Markov random field over G_1 is a Markov random field over G_2.

Example 4.1. For all graphs G and T such that $G \to T$, the uniform distribution on $\mathsf{Hom}(G,T)$, viewed as an element of $\mathsf{Dist}((V_T)^{V_G})$, is a Markov random field over G with entropy $\log \hom(G,T)$.

The next lemma gives a mechanism for constructing one Markov random field from another.

Lemma 4.2 (Pullback of a MRF)**.** *Let φ be a homomorphism from a chordal graph F to a graph G. Then for every $X \in \mathsf{MRF}(G,\Omega)$ there exists a unique $\widetilde{X} \in \mathsf{MRF}(F,\Omega)$ (called the* pullback of X along *φ) such that for every clique $C \in \mathsf{Cliques}(F)$, marginal distributions $\langle \widetilde{X}_c : c \in C\rangle$ and $\langle X_{\varphi(c)} : c \in C\rangle$ are identical. Moreover, if $\Omega = V_T$ where T is graph such that $\mathsf{Supp}(X) \subseteq \mathsf{Hom}(G,T)$, then $\mathsf{Supp}(\widetilde{X}) \subseteq \mathsf{Hom}(F,T)$.*

We already saw pullbacks of Markov random fields in action when we computed $\mathsf{HDE}(\mathit{Vee}, \vec{C}_3)$ in §1.2.

Proof Sketch. We can construct $\widetilde{X}$ according to the following procedure. Fix an arbitrary elimination ordering $v_1,\dots,v_n$ of F (so that v_j is an eliminable vertex of $F|_{\{v_1,\dots,v_j\}}$ for all $j \in [n]$). We now pick values for $\widetilde{X}_{v_1},\dots,\widetilde{X}_{v_n}$ (i.e., the coordinates of joint distribution $\widetilde{X} = (\widetilde{X}_v)_{v\in F} \in \mathsf{Dist}(\Omega^{V_F})$) in order. Assuming values $\widetilde{X}_{v_1},\dots,\widetilde{X}_{v_{j-1}}$ have been picked, we next pick $\widetilde{X}_{v_j}$ according to the distribution $X_{\varphi(v_j)}$ conditioned on $X_{\varphi(v_i)} = \widetilde{X}_{v_i}$ for $i = 1,\dots,j-1$.

One can show that the resulting distribution $\widetilde{X}$ is a Markov random field over F. Indeed, it is the unique Markov random field meeting the conditions of the lemma; in particular $\widetilde{X}$ is independent of the particular elimination ordering $v_1,\dots,v_n$ of F. In the event that $\Omega = V_T$ where T is graph such that $\mathsf{Supp}(X) \subseteq \mathsf{Hom}(G,T)$, it is easy to show that every point of $(V_T)^{V_F}$ in the support of $\widetilde{X}$ is a homomorphism in $\mathsf{Hom}(F,T)$. □

5 Proof of Theorem 3.1 (HDE Lower Bound for Chordal F)

Suppose F is chordal and $\mathsf{Hom}(F,G)$ is nonempty. Let T be a graph such that $\hom(G,T) \geqslant 2$. Let $X \in \mathsf{Dist}((V_T)^{V_G})$ be the uniform distribution on $\mathsf{Hom}(G,T)$ (so $X \in \mathsf{MRF}(G)$, see Example 4.1). Let $h_X : \wp(V_G) \longrightarrow [0,1]$ be the normalized entropy function of X and note that $h_X \in \mathcal{P}(G)$ and

$$h_X(A) = \mathbb{H}(X_A)/\log\hom(G,T).$$

For each homomorphism $\varphi \in \mathsf{Hom}(F,G)$, let $Y^\varphi \in \mathsf{MRF}(F,V_T)$ be the pullback of X along φ, as described in Lemma 4.2. We have $\mathsf{Supp}(Y^\varphi) \subseteq \mathsf{Hom}(F,T)$ and hence $\mathbb{H}(Y^\varphi) \leqslant \log\hom(F,T)$.

By equation (3) we have the following identity (independent of the graph T):

$$\mathbb{H}(Y^\varphi) = \sum_{S\subseteq \mathsf{MaxCliques}(F)} -(-1)^{|S|}\mathbb{H}(X_{\varphi(\cap S)}) = \sum_{S\subseteq \mathsf{MaxCliques}(F)} -(-1)^{|S|} h_X(\varphi(\textstyle\bigcap S))\mathbb{H}(X).$$

It follows that

$$\log\hom(F,T) \geqslant \max_{\varphi\in\mathsf{Hom}(F,G)} \sum_{S\subseteq\mathsf{MaxCliques}(F)} -(-1)^{|S|} h_X(\varphi(\textstyle\bigcap S))\log\hom(G,T).$$

Since this inequality holds for all graphs T such that $\hom(G,T) \geqslant 2$, we have

$$\begin{aligned}\mathsf{HDE}(F,G) &= \inf_{T\,:\,\hom(G,T)\geqslant 2} \frac{\log\hom(F,T)}{\log\hom(G,T)} \quad \text{(by (2))}\\ &\geqslant \inf_{T\,:\,\hom(G,T)\geqslant 2}\ \max_{\varphi\in\mathsf{Hom}(F,G)} \sum_{S\subseteq\mathsf{MaxCliques}(F)} -(-1)^{|S|} h_X(\varphi(\textstyle\bigcap S)).\end{aligned}$$

Since $h_X \in \mathcal{P}(G)$ for all T, we get the desired result that

$$\mathsf{HDE}(F,G) \geqslant \min_{p \in \mathcal{P}(G)} \max_{\varphi \in \mathsf{Hom}(F,G)} \sum_{S \subseteq \mathsf{MaxCliques}(F)} -(-1)^{|S|} p(\varphi(\textstyle\bigcap S)).$$

6 Proof of Theorem 3.2 (HDE Upper Bound)

Fix a graph G and a function $q \in \mathcal{Q}(G)$. That is, let q be a function from $\wp(V_G)$ to $[0,1]$ such that $q(\emptyset) = 0$ and $\sum_{A \subseteq V_G} q(A) \cdot \mathsf{CC}(G|_A) = 1$.

We define a sequence $(T_n)_{n \geqslant 1}$ of "target" graphs as follows. Vertices of T_n are all pairs (x,i) where $x \in V_G$ and $i \in \mathbb{N}^{\{A \subseteq V_G : x \in A\}}$ is a function from $\{A \subseteq V_G : x \in A\}$ to $\mathbb{N}$ which satisfies $i(A) < n^{q(A)}$. There is an edge in T_n from vertex (x,i) to vertex (y,j) if and only if $(x,y) \in E_G$ and $i(A) = j(A)$ for all $\{x,y\} \subseteq A \subseteq V_G$.

Let π_n denote the homomorphism from T_n to G defined by $\pi_n((x,i)) = x$. Let F be a graph and suppose φ is a homomorphism from F to G. We denote by $\mathsf{Hom}_\varphi(F,T_n)$ the set of homomorphisms $\psi : F \to T_n$ such that $\pi_n \circ \psi = \varphi$, i.e., the following diagram commutes:

$$\begin{tikzcd} & T_n \arrow[d, "\pi_n"] \\ F \arrow[ur, "\psi"] \arrow[r, "\varphi"] & G \end{tikzcd}$$

Let $\hom_\varphi(F,T_n) = |\mathsf{Hom}_\varphi(F,T_n)|$ and note that

$$\hom(F,T_n) = \sum_{\varphi \in \mathsf{Hom}(F,G)} \hom_\varphi(F,T_n). \tag{4}$$

Lemma 6.1. $\displaystyle\lim_{n\to\infty} \log_n \hom_\varphi(F,T_n) = \sum_{A \subseteq V_G} q(A) \cdot \mathsf{CC}(F|_{\varphi^{-1}(A)})$.

Proof. Let $\psi \in \mathsf{Hom}_\varphi(F,T_n)$. Each vertex $u \in V_F$ is mapped under ψ to a pair $(\varphi(u), i_u)$ for some $i_u \in \mathbb{N}^{\{A \subseteq V_G : \varphi(u) \in A\}}$ subject to $i_u(A) < n^{q(A)}$. The family of functions $(i_u)_{u \in V_F}$ is further subject to the constraint that $i_u(A) = i_v(A)$ for all $u, v \in V_F$ and $\{\varphi(u), \varphi(v)\} \subseteq A \subseteq V_G$ such that u and v lie in the same connected component of $F|_{\varphi^{-1}(A)}$. To see this, consider an undirected path in $F|_{\varphi^{-1}(A)}$ from u to v, i.e., a sequence $u = w_0, w_1, w_2, \dots, w_k = v$ such that $(w_{\ell-1}, w_\ell)$ or $(w_\ell, w_{\ell-1})$ is an edge in $F|_{\varphi^{-1}(A)}$ for every $\ell \in \{1,\dots,k\}$. Suppose $\{\varphi(u),\varphi(v)\} \subseteq A \subseteq V_G$ and u, v lie in the same connected component of $F|_{\varphi^{-1}(A)}$. Then clearly $\{\varphi(w_{\ell-1}), \varphi(w_\ell)\} \subseteq A$ for all $\ell \in \{1,\dots,k\}$. Since $(w_{\ell-1}, w_\ell)$ or $(w_\ell, w_{\ell-1})$ is an edge in F and ψ is a homomorphism from F to T_n, we have that $(\psi(w_{\ell-1}), \psi(w_\ell))$ or $(\psi(w_\ell), \psi(w_{\ell-1}))$ is an edge in T_n. It follows that $i_{\varphi(w_{\ell-1})}(B) = i_{\varphi(w_\ell)}(B)$ for all $\{\varphi(w_{\ell-1}), \varphi(w_\ell)\} \subseteq B \subseteq V_G$. In particular, we have $i_{\varphi(w_{\ell-1})}(A) = i_{\varphi(w_\ell)}(A)$. Therefore $i_u(A) = i_{w_0}(A) = \dots = i_{w_k}(A) = i_v(A)$.

Conversely, every family of functions $\langle j_u \in \mathbb{N}^{\{A \subseteq V_G : \varphi(u) \in A\}} : u \in V_F \rangle$ subject to $j_u(A) < n^{q(A)}$ and $j_u(A) = j_v(A)$ for all $u, v \in V_F$ and $\{\varphi(u), \varphi(v)\} \subseteq A \subseteq V_G$ such that u and v lie in the same connected component of $F|_{\varphi^{-1}(A)}$, determines a distinct homomorphism in $\mathsf{Hom}_\varphi(F,T_n)$. Thus, $\hom_\varphi(F,T_n)$ equals the number of such families $(j_u)_{u \in V_F}$. This is precisely $\prod_{A \subseteq V_G} \lceil n^{q(A) \cdot \mathsf{CC}(F|_{\varphi^{-1}(A)})} \rceil$, since for each $A \subseteq V_G$ and each connected component U of $F|_{\varphi^{-1}(A)}$, we have an independent choice of numbers $m_{A,U} \in \{0, \dots, \lceil n^{q(A)} \rceil - 1\}$ such that $j_u(A) = m_{A,U}$ for all $u \in U$. Taking logarithms in base n, we get the statement of the lemma. $\square$

Corollary 6.2. $\lim_{n\to\infty} \log_n \hom(F,T_n) = \max_{\varphi\in\mathsf{Hom}(F,G)} \sum_{A\subseteq V_G} q(A)\cdot \mathsf{CC}(F|_{\varphi^{-1}(A)})$.

This corollary follows immediately from (4) and Lemma 6.1. We are ready to prove Theorem 3.2.

Proof of Theorem 3.2. Suppose $F \to G$. For $q \in \mathcal{Q}(G)$, let $(T_n)_{n\geqslant 1}$ be the sequence of "target" graphs as above. By Corollary 6.2 (applied to G), we have

$$\lim_{n\to\infty} \log_n \hom(G,T_n) = \max_{\varphi\in\mathsf{Hom}(G,G)} \sum_{A\subseteq V_G} q(A)\cdot \mathsf{CC}(G|\varphi^{-1}(A)) \geqslant \sum_{A\subseteq V_G} q(A)\cdot \mathsf{CC}(G|_A) = 1$$

where the middle inequality is obtained by taking φ to be the identity homomorphism on G.

We now have

$$\mathsf{HDE}(F,G) \overset{(2)}{\leqslant} \lim_{n\to\infty} \frac{\log_n \hom(F,T_n)}{\log_n \hom(G,T_n)} \leqslant \lim_{n\to\infty} \log_n \hom(F,T_n) = \max_{\varphi\in\mathsf{Hom}(F,G)} \sum_{A\subseteq V_G} q(A)\cdot \mathsf{CC}(F|_{\varphi^{-1}(A)})$$

where the last equality is by Corollary 6.2. Since this inequality holds for all $q \in \mathcal{Q}(G)$, it follows that

$$\mathsf{HDE}(F,G) \leqslant \min_{q\in\mathcal{Q}(G)} \max_{\varphi\in\mathsf{Hom}(F,G)} \sum_{A\subseteq V_G} q(A)\cdot \mathsf{CC}(F|_{\varphi^{-1}(A)}). \qquad \square$$

7 Proof of Theorem 3.3 (HDE of Chordal F and Series-Parallel G)

Suppose F is chordal and G is series-parallel and $F \to G$. The HDE lower bound of Theorem 3.1 states

$$\mathsf{HDE}(F,G) \geqslant \min_{p\in\mathcal{P}(G)} \max_{\varphi\in\mathsf{Hom}(F,G)} \sum_{S\subseteq \mathsf{MaxCliques}(F)} -(-1)^{|S|}\cdot p(\varphi(\textstyle\bigcap S)).$$

Let p be an arbitrary function in $\mathcal{P}(G)$. To prove Theorem 3.3 (i.e., to prove this inequality is tight), we construct a sequence of graphs T_n satisfying

$$\lim_{n\to\infty} \log_n \hom(G,T_n) \geqslant 1, \tag{5}$$

$$\lim_{n\to\infty} \log_n \hom(F,T_n) \leqslant \max_{\varphi\in\mathsf{Hom}(F,G)} \sum_{S\subseteq \mathsf{MaxCliques}(F)} -(-1)^{|S|} p(\varphi(\textstyle\bigcap S)). \tag{6}$$

Tightness of the above HDE lower bound then follows thanks to (2).

To simplify matters, we first consider the special case that G is chordal. (Since G is chordal and series-parallel, it has clique number $\leqslant 3$, i.e., G is a 2-tree.) After proving Theorem 3.3 in this special case, we give the argument for general series-parallel G in Section 7.4.

We construct $T = T_n$ in two stages. For every $A \in \mathsf{MaxCliques}(G)$, we construct a graph T_A together with a homomorphism $\pi_A : T_A \longrightarrow K_A$ (the complete graph on A, viewed as a subgraph of G). We then patch together (via a randomized gluing procedure) the various graphs T_A into a graph T together with a homomorphism $\pi : T \longrightarrow G$. (This indexing over maximal cliques in the chordal graph G is essential to defining the gluing procedure in a consistent fashion.)

For $a, b, c \in V_G$, we write $p(a), p(ab), p(abc)$ for $p(\{a\}), p(\{a,b\}), p(\{a,b,c\})$ respectively. For $A \subseteq V_G$, we treat $n^{p(A)}$ as integers (by rounding), mindful to preserve identities such as $n^{p(a)+p(bc)} = n^{p(a)}n^{p(bc)}$. Because we are ultimately interested in asymptotics in log base n, this kind of rounding presents no difficulties.

7.1 Construction of T_A

Consider any $A \in \mathsf{MaxCliques}(G)$ and note that $|A| \in \{1,2,3\}$.

If $|A| = 1$ (say $A = \{a\}$), then T_A is the empty (edgeless) graph on $n^{p(\{a\})}$ vertices and π_A maps all vertices of T_A to a.

Now suppose $|A| = 2$ (say $A = \{a,b\}$). Letting

$$\alpha = n^{p(\{a\})}, \quad \beta = n^{p(\{b\})}, \quad \gamma = n^{p(\{a\})+p(\{b\})-p(\{a,b\})} \tag{7}$$

(note that $\gamma \geqslant 1$ by submodularity of p), T_A is the graph $\gamma{\cdot}K_{\alpha,\beta}$ (i.e., γ disjoint copies of the complete bipartite graph $K_{\alpha,\beta}$) and $\pi_A \in \mathsf{Hom}(T_A, K_A)$ maps the two parts of each $K_{\alpha,\beta}$ to vertices a and b of K_A (i.e., the α-size part to a and the β-size part to b).

We now examine the nontrivial case when $|A| = 3$ (say $A = \{a,b,c\}$). Consider the restriction of p to $\wp(A)$. So long as $p(A) > 0$, the normalized function $\frac{p}{p(A)} \restriction \wp(A)$ is K_A-polymatroidal (if $p(A) = 0$, then $p \restriction \wp(A)$ is identically zero). By Example 2.4, it follows that $p \restriction \wp(A)$ is a nonnegative linear combination of functions $f_a, f_b, f_c, f_{ab}, f_{ac}, f_{bc}, f_{abc}$ and f_{RS}. That is,

$$p \restriction \wp(A) = \sum_{i \in \{a,b,c,ab,ac,bc,abc,\mathrm{RS}\}} \lambda_i f_i \text{ for some } \lambda_i \geqslant 0.$$

(We will harmlessly treat n^{λ_i} as integers.) Note the identities:

$$\begin{aligned} p(a) &= \lambda_a + \lambda_{ab} + \lambda_{ac} + \lambda_{abc} + \tfrac{1}{2}\lambda_{\mathrm{RS}}, \\ p(ab) &= \lambda_a + \lambda_b + \lambda_{ab} + \lambda_{ac} + \lambda_{bc} + \lambda_{abc} + \tfrac{1}{2}\lambda_{\mathrm{RS}}, \\ p(abc) &= \lambda_a + \lambda_b + \lambda_c + \lambda_{ab} + \lambda_{ac} + \lambda_{bc} + \lambda_{abc} + \tfrac{1}{2}\lambda_{\mathrm{RS}}. \end{aligned} \tag{8}$$

For each $i \in \{a,b,c,ab,ac,bc,abc,\mathrm{RS}\}$, we will construct a graph $T_{A,i}$ and a homomorphism $\pi_{A,i} : T_{A,i} \longrightarrow K_A$. Once we have defined these, we obtain T_A as the *fibered product* of graphs $T_{A,i}$:

- the vertices of T_A are the elements $(v_i) \in \prod_i T_{A,i}$ such that $\pi_{A,i}(v_i) = \pi_{A,j}(v_j)$ for all $i,j \in \{a,b,c,ab,ac,bc,abc,\mathrm{RS}\}$, and
- there is an edge between vertices (v_i) and (w_i) of T_A if and only if there is an edge between v_i and w_i in $T_{A,i}$ for every $i \in \{a,b,c,ab,ac,bc,abc,\mathrm{RS}\}$.

The homomorphism $\pi_A : T_A \longrightarrow K_A$ is defined in the obvious way:

- $\pi_A((v_i))$ equals the common value of $\pi_{A,i}(v_i)$.

We now define $T_{A,i}$ and $\pi_{A,i}$ for the various $i \in \{a,b,c,ab,ac,bc,abc,\mathrm{RS}\}$. In all cases, after defining $T_{A,i}$, the homomorphism $\pi_{A,i}$ will be obvious. Also, the definitions of $T_{A,b}$ and $T_{A,c}$ will be obvious after stating the definition of $T_{A,a}$, so we include only the cases $i \in \{a, ab, abc, \mathrm{RS}\}$.

- $T_{A,a}$ has vertex set $(\{a\} \times [n^{\lambda_a}]) \cup \{b,c\}$ and edges $\{b,c\}$ and $\{(a,i),b\}$ and $\{(a,i),c\}$ for all $i \in [n^{\lambda_a}]$.
- $T_{A,ab}$ has vertex set $(\{a,b\} \times [n^{\lambda_{ab}}]) \cup \{c\}$ and edges $\{(a,i),(b,i)\}$ and $\{(a,i),c\}$ and $\{(b,i),c\}$ for all $i \in [n^{\lambda_{ab}}]$.
- $T_{A,abc}$ has vertex set $\{a,b,c\} \times [n^{\lambda_{abc}}]$ and edges $\{(a,i),(b,i)\}$ and $\{(a,i),(c,i)\}$ and $\{(b,i),(c,i)\}$ for all $i \in [n^{\lambda_{abc}}]$.
- If $\lambda_{\mathrm{RS}} = 0$, then $T_{A,\mathrm{RS}} = K_A$ and π_A is the identity function on A.

To define the remaining graph $T_{A,\mathrm{RS}}$ when $\lambda_{\mathrm{RS}} > 0$, we use a result of Ruzsa and Szemeredi [RS78].

Theorem 7.1 (Ruzsa-Szemeredi [RS78]). *For all $m \in \mathbb{N}$, there exists a tripartite graph $H(m)$ in which:*

(i) *each part has size m,*

(ii) *there are $m^{2-o(1)}$ triangles, and*

(iii) *every edge is contained in exactly one triangle.*

(This is not the usual statement of the Ruzsa-Szemeredi result. However, it is easily seen to be equivalent to the usual statement that there exists a bipartite graph with parts of size m whose edge set is the disjoint union of $m^{1-o(1)}$ induced matchings of size at least $m^{1-o(1)}$.)

Using Theorem 7.1, we define $T_{A,\mathrm{RS}}$ in the remaining case:

- If $\lambda_{\mathrm{RS}} > 0$, let $T_{A,\mathrm{RS}}$ be the graph $H(n^{\frac{1}{2}\lambda_{\mathrm{RS}}})$ of Theorem 7.1 and let $\pi_{A,\mathrm{RS}} \in \mathsf{Hom}(T_{A,\mathrm{RS}}, K_A)$ be any function mapping the three parts to a, b and c.

Recalling the definition of T_A (as a fibered product of graphs $T_{A,i}$), it is easy to check using equations (8) that the graph T_A satisfies:

$$\begin{aligned} |\{\text{vertices of } T_A \text{ which map to } a \text{ under } \pi_A\}| &= n^{p(a)}, \\ |\{\text{edges of } T_A \text{ which map to } \{a,b\} \text{ under } \pi_A\}| &= n^{p(ab)-o(1)}, \\ |\{\text{triangles in } T_A\}| &= n^{p(abc)-o(1)}. \end{aligned}$$

Moreover, the $o(1)$ terms disappear whenever $\lambda_{\mathrm{RS}} = 0$.

7.2 Gluing Procedure

We now describe the randomized procedure for gluing together the various graphs T_A and homomorphisms $\pi_A : T_A \longrightarrow K_A$ into a single graph T and homomorphism $\pi : T_A \longrightarrow G$. It is enough to describe the procedure for gluing a pair of graphs T_A and T_B for $A, B \in \mathsf{MaxCliques}(G)$: there is an obvious way of simultaneously and consistently carrying out all pairwise gluings to obtain T and π (relying on the chordality of G).

Let $A, B \in \mathsf{MaxCliques}(G)$. There are three gluing procedures to consider, depending on $|A \cap B| \in \{0, 1, 2\}$. In the simplest case that $A \cap B = \emptyset$, the gluing of T_A and T_B is just the disjoint union $T_A \uplus T_B$ and gluing of homomorphisms π_A and π_B is obvious.

Next suppose that $|A \cap B| = 1$ (say $A \cap B = \{a\}$). Note that $|\pi_A^{-1}(a)| = |\pi_B^{-1}(a)| = n^{p(a)}$. The gluing of T_A and T_B is defined by starting with the disjoint union $T_A \uplus T_B$ and identifying pairs of vertices in $\pi_A^{-1}(a) \times \pi_B^{-1}(a)$ under a uniformly choosen random bijection between sets $\pi_A^{-1}(a)$ and $\pi_B^{-1}(a)$.

Finally, suppose that $|A \cap B| = 2$ (say $A \cap B = \{a, b\}$). In this case, it must happen that $|A| = |B| = 3$. Define α, β, γ again by equation (7) and consider the graph $\gamma{\cdot}K_{\alpha,\beta}$. We claim that bipartite graphs $T_A|_{\pi_A^{-1}(\{a,b\})}$ and $T_B|_{\pi_B^{-1}(\{a,b\})}$ both look like $\gamma{\cdot}K_{\alpha,\beta}$ after deleting an $n^{-o(1)}$-fraction of edges from the latter. (The proof of Claim 7.2, below, follows easily from definitions.)

Claim 7.2. *There exist homomorphisms $\xi_A : T_A|_{\pi_A^{-1}(\{a,b\})} \longrightarrow \gamma{\cdot}K_{\alpha,\beta}$ and $\xi_B : T_A|_{\pi_B^{-1}(\{a,b\})} \longrightarrow \gamma{\cdot}K_{\alpha,\beta}$ such that*

- *ξ_A and ξ_B are bijections (between vertex sets), and*
- *ξ_A maps $\pi_A^{-1}(a)$ to the α-side of $\gamma{\cdot}K_{\alpha,\beta}$ and $\pi_A^{-1}(b)$ to the β-side of $\gamma{\cdot}K_{\alpha,\beta}$, and similarly for ξ_B.*

Moreover, $T_A|_{\pi_A^{-1}(\{a,b\})}$ *and* $T_B|_{\pi_B^{-1}(\{a,b\})}$ *both have at least* $n^{\alpha+\beta+\gamma-o(1)}$ *edges (thus, these graphs may be obtained from* $\gamma{\cdot}K_{\alpha,\beta}$ *by deleting an* $n^{-o(1)}$*-fraction of edges).*

After fixing arbitrary ξ_A and ξ_B, the gluing procedure works as follows. We pick a uniform random automorphism Ψ of $\gamma{\cdot}K_{\alpha,\beta}$ (i.e., an element of the group $(S_\alpha \times S_\beta) \ltimes S_\gamma$). The function $\xi_B^{-1} \circ \Psi \circ \xi_A$ is a bijection of sets $\pi_A^{-1}(\{a,b\})$ and $\pi_B^{-1}(\{a,b\})$. Starting from the disjoint union of T_A and T_B, we identify pairs of vertices under this bijection. Finally, we keep edges between pairs of identified vertices if and only if edges existed between these vertices in both T_A and T_B. (Intuitively, we randomly overlap T_A and T_B within the confines of $\gamma{\cdot}K_{\alpha,\beta}$ and keep only the edges which occur in both T_A and T_B.)

Having defined randomized gluings for pairs of graphs T_A and T_B, suffice it to say that these pairwise gluing can without difficulty be carried out simultaneously and consistently over all $A \in \mathsf{MaxCliques}(G)$ to obtain the graph T and homomorphism $\pi : T \longrightarrow G$ (chordality of G is crucial here).

7.3 Counting Homomorphisms from F and G

Now that we have defined the sequence of graphs T_n and homomorphisms $\pi_n : T_n \longrightarrow G$, it remains to prove inequalities (5) and (6). Both inequalities follow from the following claim.

Claim 7.3. *If* H *is a chordal graph and* $\varphi \in \mathsf{Hom}(H,G)$*, then*

$$\log_n |\{\theta \in \mathsf{Hom}(H,T_n) : \pi_n \circ \theta = \varphi\}| = \sum_{S \subseteq \mathsf{MaxCliques}(H)} -(-1)^{|S|} p(\varphi(\textstyle\bigcap S)) - o(1).$$

Before proving of Claim 7.3, let's see how it implies inequalities (5) and (6). To prove (5), we take $H = G$ and $\varphi = \mathrm{id}_{V_G}$ (the identity map on V_G viewed a homomorphism $G \longrightarrow G$) in Claim 7.3 and see

$$\begin{aligned}\log_n \hom(G,T_n) &\geqslant \log_n |\{\theta \in \mathsf{Hom}(G,T_n) : \pi_n \circ \theta = \mathrm{id}_{V_G}\}| \\ &= \sum_{S \subseteq \mathsf{MaxCliques}(G)} -(-1)^{|S|} p(\textstyle\bigcap S) - o(1) = 1 - o(1) \quad \text{(by Lemma 2.5).}\end{aligned}$$

Inequality (6) is immediate from Claim 7.3 taking $H = F$:

$$\begin{aligned}\lim_{n\to\infty} \log_n \hom(F,T_n) &= \lim_{n\to\infty} \max_{\varphi \in \mathsf{Hom}(F,G)} \log_n |\{\theta \in \mathsf{Hom}(F,T_n) : \pi_n \circ \theta = \varphi\}| \quad (\text{as } \hom(F,T_n) \xrightarrow{n\to\infty} \infty) \\ &= \sum_{S \subseteq \mathsf{MaxCliques}(H)} -(-1)^{|S|} p(\varphi(\textstyle\bigcap S)).\end{aligned}$$

Now for the proof of this claim:

Proof of Claim 7.3. We define a supergraph T^* of T as follows. For each $A \in \mathsf{MaxCliques}(G)$, we define a supergraph T_A^* of T_A and apply the same gluing procedure. If $|A| \leqslant 2$, let $T_A^* = T_A$. If $|A| = 3$ (say $A = \{a,b,c\}$), recall that T_A is the fibred product of graphs $T_{A,a}, \ldots, T_{A,abc}$ and $T_{A,\mathrm{RS}}$; let T_A^* be the fibred product of graphs $T_{A,a}, \ldots, T_{A,abc}$ and $T_{A,\mathrm{RS}}^*$ where $T_{A,\mathrm{RS}}^*$ is the complete tripartite graph with all parts of size $n^{\frac{1}{2}\lambda_{\mathrm{RS}}(A)}$. Viewing $T_{A,\mathrm{RS}}$ as a subgraph of $T_{A,\mathrm{RS}}^*$ (with the same vertex set) and apply the same gluing procedure (i.e., with the same randomization), we view T as a subgraph of T^* (with the same vertex set). It now suffices to prove the following:

$$(9)\qquad \log_n |\{\theta \in \mathsf{Hom}(H, T_n^*) : \pi_n \circ \theta = \varphi\}| = \sum_{S \subseteq \mathsf{MaxCliques}(H)} -(-1)^{|S|} p(\varphi(\textstyle\bigcap S)) + \sum_{A \in \mathsf{MaxCliques}(G)\,:\,|A|=3} \tfrac{1}{2}\lambda_{\mathrm{RS}}(A) \cdot |\{A' \in \mathsf{MaxCliques}(H) : \varphi(A') = A\}|,$$

$$(10)\qquad \log_n \Pr_{\theta \in \mathsf{Hom}(H,T_n^*)}[\theta \in \mathsf{Hom}(H, T_n)] = - \sum_{A \in \mathsf{MaxCliques}(G)\,:\,|A|=3} \tfrac{1}{2}\lambda_{\mathrm{RS}}(A) \cdot |\{A' \in \mathsf{MaxCliques}(H) : \varphi(A') = A\}| - o(1).$$

We first give the argument for equation (9). Note the following:

- for every edge (a, b) in G and every $a' \in \pi_n^{-1}(a)$,
$$|\{b' \in \pi_n^{-1}(b) : (a', b') \text{ is an edge in } T_n^*\}| = n^{p(ab)-p(a)},$$
- for every triangle (a, b, c) in G and every $a' \in \pi_n^{-1}(a)$ and $b' \in \pi_n^{-1}(b)$ such that (a', b') is an edge in T_n^*,
$$|\{c' \in \pi_n^{-1}(c) : (a', b', c') \text{ is a triangle in } T_n^*\}| = n^{p(abc)-p(ab)+\frac{1}{2}\lambda_{\mathrm{RS}}(abc)}.$$

It follows that, if is $v_1, \ldots, v_n$ an elimination ordering for H, then

$$\log_n |\{\theta \in \mathsf{Hom}(H, T_n^*) : \pi_n \circ \theta = \varphi\}| = \sum_{i=1}^{n} p\big(\varphi(\{\text{neighbors of } v_i \text{ among } v_1, \ldots, v_{i-1}\} \cup \{v_i\})\big) - p\big(\varphi(\{\text{neighbors of } v_i \text{ among } v_1, \ldots, v_{i-1}\})\big) + \sum_{A \in \mathsf{MaxCliques}(G)\,:\,|A|=3} \tfrac{1}{2}\lambda_{\mathrm{RS}}(A) \cdot |\{A' \in \mathsf{MaxCliques}(H) : \varphi(A') = A\}|.$$

Equation (9) now follows using Lemma 2.5.

For equation (10), notice that a triangle (a', b', c') over (a, b, c) in T_n^* is a triangle in T_n with probability $n^{-\lambda_{\mathrm{RS}}(abc)-o(1)}$. Now consider a uniform random homomorphism $\theta \in \mathsf{Hom}(H, T_n^*)$. For an edge (x, y) in H, consider the vertices $z_1, \ldots, z_m$ such that (x, y, z_j) are triangles in H. The key observation (using chordality of H) is that events $\{(\theta(x), \theta(y), \theta(z_j)) \text{ is a triangle in } T_n\}_{j=1,\ldots,m}$ are independent conditioned on $\theta(x)$ and $\theta(y)$. By expanding the probability that $\theta \in \mathsf{Hom}(H, T_n)$ conditionally along an elimination ordering, we see that $\theta \notin \mathsf{Hom}(H, T_n)$ with probability $\prod_{\text{triangles } (x,\,y,\,z) \text{ in } H} n^{-\lambda_{\mathrm{RS}}(\theta(x)\theta(y)\theta(z))-o(1)}$, which proves (10) and completes the proof of Claim 7.3. □

7.4 Series-Parallel G

Finally, we prove the theorem for the case when G is series-parallel (but not necessarily chordal). Recall that for every series-parallel graph G, there exists a 2-tree $\widetilde{G}$ (i.e., a K_4-free chordal graph) such that $V_G = V_{\widetilde{G}}$ and $E_G \subseteq E_{\widetilde{G}}$. Fix any such $\widetilde{G}$.

Consider any $p \in \mathcal{P}(G)$. Note that $\mathcal{P}(G) \subseteq \mathcal{P}(\widetilde{G})$ (i.e., any normalized G-polymatroidal function is also normalized $\widetilde{G}$-polymatroidal). Therefore, we can construct graphs $\widetilde{T}_n$ with homomorphisms π_n :

$\widetilde{T}_n \longrightarrow \widetilde{G}$ such that (by Claim 7.3 applied to $\widetilde{G}$ and $\widetilde{T}_n$) for every chordal graph H and $\varphi \in \mathsf{Hom}(H, \widetilde{G})$,

$$\log_n |\{\theta \in \mathsf{Hom}(H, \widetilde{T}_n) : \pi_n \circ \theta = \varphi\}| = \sum_{S \subseteq \mathsf{MaxCliques}(H)} -(-1)^{|S|} p(\varphi(\bigcap S)) - o(1). \tag{11}$$

Let T_n be the subgraph of $\widetilde{T}_n$ which has the same vertices, but where we keep an edge (v, w) from $\widetilde{T}_n$ if and only if $(\pi_n(v), \pi_n(w))$ is an edge of G. Note that π_n is a homomorphism in $\mathsf{Hom}(T_n, G)$. By (11), Claim 7.3 now holds (exactly as stated) for G and T_n. The proof of inequalities (5) and (6) then follows by the exact same argument.

8 Proof of Theorem 3.4 (HDE of P_4 and P_{4n+2})

Let $P_{4n+2} = (V, E)$ where $V = \{0, 1, \ldots, 4n+1\}$ and $E = \big\{\{0,1\}, \{1,2\}, \ldots, \{4n, 4n+1\}\big\}$. Define function $f : V \longrightarrow \mathbb{N}$ as follows:

- $f(0) = f(4n+1) = 2n+1$,
- $f(4k+1) = f(4k+3) = 2k+1$ for $k \in \{0, \ldots, n-1\}$,
- $f(4k+2) = f(4k+4) = 2n-2k-1$ for $k \in \{0, \ldots, n-1\}$.

For every $N \in \mathbb{N}$, we define a random graph $T_N = (V_N, E_N)$ as follows. Let

$$V_N = \big\{(v, i) : v \in V,\ i \in \{1, \ldots, \lceil N^{f(v)} \rceil\}\big\}.$$

Independently for all $(v, i), (w, j) \in V_N$, place an edge with probability

$$\Pr\big[\{(v,i),(w,j)\} \in E_N\big] = \begin{cases} \frac{1}{N} & \text{if } \{v, w\} = \{4k, 4k+1\} \text{ where } k \in \{0, \ldots, n-1\}, \\ 1 & \text{if } \{v, w\} = \{4k+r, 4k+r+1\} \text{ where } k \in \{0, \ldots, n-1\} \text{ and } r \in \{1, 2, 3\}, \\ 0 & \text{otherwise.} \end{cases}$$

It holds with high probability that

$$\hom(P_{4n+2}, T_N) \geqslant N^{4n^2+3n+1-o(1)}.$$

It also holds with high probability (by inspection of the various homomorphisms from P_4 to P_{4n+2}) that

$$\hom(P_4, T_N) \leqslant N^{4n+1+o(1)}.$$

Therefore,

$$\mathsf{HDE}(P_4, P_{4n+2}) \leqslant \frac{4n+1}{4n^2+3n+1}.$$

We now prove the opposite inequality. We will represent homomorphisms $P_4 \longrightarrow P_{4n+2}$ by 4-tuples $\langle i_1, i_2, i_3, i_4 \rangle \in V^4$. Define a function $w : \mathsf{Hom}(P_4, P_{4n+2}) \longrightarrow \mathbb{N}$ as follows:

$$\begin{aligned} w(\langle 4k, 4k+1, 4k, 4k+1 \rangle) &= 1 && \text{for } k \in \{0, \ldots, n\}, \\ w(\langle 4k, 4k+1, 4k+2, 4k+1 \rangle) &= 1 && \text{for } k \in \{0, \ldots, n-1\}, \\ w(\langle 4(n-k)+1, 4(n-k), 4(n-k)-1, 4(n-k) \rangle) &= 1 && \text{for } k \in \{0, \ldots, n-1\}, \\ w(\langle 4k+2, 4k+3, 4k+4, 4k+5 \rangle) &= 4k+2 && \text{for } k \in \{0, \ldots, n-1\}, \\ w(\langle 4(n-k)+1, 4(n-k), 4(n-k)-1, 4(n-k)-2 \rangle) &= 4k+2 && \text{for } k \in \{0, \ldots, n-1\}, \end{aligned}$$

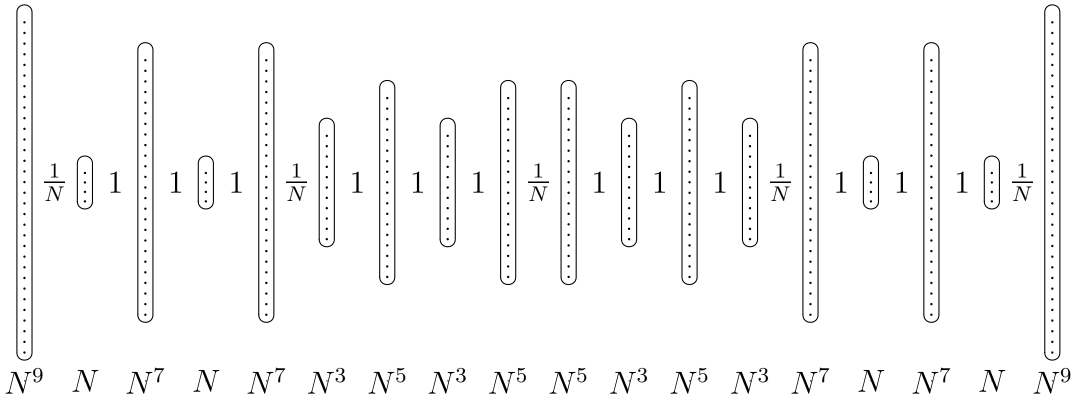


Figure 1: The random graph T_N when $n = 4$ (drawn to logscale height). The number (1 or $\frac{1}{n}$) in-between partitions of the vertex set indicates the probability of an edge.

and let $w(\varphi) = 0$ for all other homomorphisms $\varphi \in \mathsf{Hom}(P_4, P_{4n+2})$. Note that

$$\sum_{\varphi \in \mathsf{Hom}(P_4, P_{4n+2})} w(\varphi) = 4n^2 + 3n + 1.$$

Fix any target graph T with at least one undirected edge. Let $X \in \mathsf{Dist}((V_T)^{V_G})$ be the uniform distribution on $\mathsf{Hom}(G, T)$. Let Φ be a random homomorphism in $\mathsf{Hom}(F, G)$ drawn according to

$$\Pr\left[\Phi = \varphi\right] = \frac{w(\varphi)}{4n^2 + 3n + 1}.$$

Let $Y^\Phi \in \mathsf{Dist}((V_T)^{V_F})$ denote the pullback of X along Φ (so in particular $\mathsf{Supp}(Y^\Phi) \subseteq \mathsf{Hom}(F, T)$).

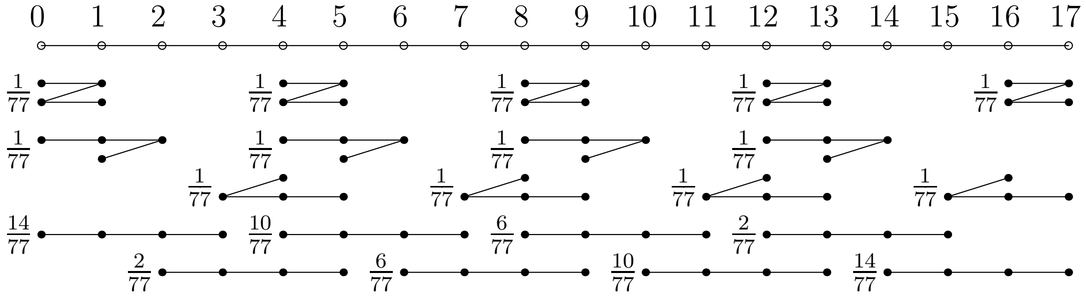


Figure 2: The distribution Φ of homomorphisms $P_4 \longrightarrow P_{4n+2}$ when $n = 4$.

By a straightforward calculation using equation (3), we have

$$
\begin{aligned}
(12)\qquad & (4n^2+3n+1)\mathbb{H}Y^\Phi \\
&= \Big(\mathbb{H}X_{\{0,1\}} - \mathbb{H}X_0\Big) + \Big(\mathbb{H}X_{\{n,n+1\}} - \mathbb{H}X_{4n+1}\Big) + \sum_{k=0}^{n}(4n+1)\mathbb{H}X_{\{4k,4k+1\}} \\
&\quad + \sum_{k=0}^{n-1}\begin{pmatrix} & (4n-4k)\mathbb{H}X_{\{4k+1,4k+2\}} \\ + & 4n\mathbb{H}X_{\{4k+2,4k+3\}} \\ + & (4k+4)\mathbb{H}X_{\{4k+3,4k+4\}}\end{pmatrix} - \begin{pmatrix} & (4n-4k)\mathbb{H}X_{4k+1} \\ + & (4n-4k-1)\mathbb{H}X_{4k+2} \\ + & (4k+3)\mathbb{H}X_{4k+3} \\ + & (4k+4)\mathbb{H}X_{4k+4}\end{pmatrix}.
\end{aligned}
$$

By monotonicity and submodularity of the entropy operator (also using the fact that $\mathbb{H}X_\emptyset = 0$), we have

$$
(13)\qquad 0 \geqslant \begin{cases} \mathbb{H}X_0 - \mathbb{H}X_{\{0,1\}}, \\ \mathbb{H}X_{4n+1} - \mathbb{H}X_{\{4n,4n+1\}}, \\ \sum_{k=0}^{n-1}(4k+1)\Big(\mathbb{H}X_{\{4k+1,4k+2\}} - \mathbb{H}X_{4k+1} - \mathbb{H}X_{4k+2}\Big), \\ \sum_{k=0}^{n-1}\mathbb{H}X_{\{4k+2,4k+3\}} - \mathbb{H}X_{4k+2} - \mathbb{H}X_{4k+3}, \\ \sum_{k=0}^{n-1}(4n-4k-3)\Big(\mathbb{H}X_{\{4k+3,4k+4\}} - \mathbb{H}X_{4k+3} - \mathbb{H}X_{4k+4}\Big). \end{cases}
$$

Adding each negative quantity in the lefthand side of equation (13) to the righthand side of equation (12), we get

$$
\begin{aligned}
(4n^2+3n+1)\mathbb{H}Y^\Phi &\geqslant (4n+1)\left(\sum_{\{v,w\}\in E}\mathbb{H}X_{\{v,w\}} - \sum_{v\in\{1,\ldots,4n\}}\mathbb{H}X_v\right) \\
&= (4n+1)\mathbb{H}X \quad \text{by (3).}
\end{aligned}
$$

It follows that $\mathsf{HDE}(P_4, P_{4n+2}) \geqslant \dfrac{4n^2+3n+1}{4n+1}$, as required.

9 Conclusion

The main open question is whether $\mathsf{HDE}(F,G)$ is computable. (This question is equivalent to decidability of the homomorphism domination problem by virtue of Lemma 2.2(3).) Theorem 3.3 shows that $\mathsf{HDE}(F,G)$ is computable in only the special case that F is chordal and G is series-parallel. Examples like $\mathsf{HDE}(\mathit{Vee}, \vec{C}_3)$ show that the homomorphism domination exponent can be tricky to compute even for very small instances. Our work also raises the finding a closed-form expression for $\mathsf{HDE}(P_m, P_n)$. So far, we only have closed expressions when m is odd or equal to 2 or 4. Besides the applications in database theory, we hope that the homomorphism domination exponent will be seen as interesting parameter in its own right.

Acknowledgements

We would like to thank Madhu Sudan, Noga Alon and Ehud Friedgut for helpful discussions.

References

[Alo81] Noga Alon. On the number of subgraphs of prescribed type of graphs with a given number of edges. *Israel J. Math.*, 38(1-2):116–130, 1981.

[BCL$^+$06] Christian Borgs, Jennifer Chayes, László Lovász, Vera T. Sós, and Katalin Vesztergombi. Counting graph homomorphisms. In *Topics in Discrete Mathematics*, volume 26 of *Algorithms Combin.*, pages 315–371. Springer, Berlin, 2006.

[FK98] Ehud Friedgut and Jeff Kahn. On the number of copies of one hypergraph in another. *Israel J. Math.*, 105:251–256, 1998.

[Fri04] Ehud Friedgut. Hypergraphs, entropy, and inequalities. *Amer. Math. Monthly*, 111(9):749–760, 2004.

[GT04] David Galvin and Prasad Tetali. On weighted graph homomorphisms. In *Graphs, morphisms and statistical physics*, volume 63 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 97–104. Amer. Math. Soc., Providence, RI, 2004.

[IR95] Yannis E. Ioannidis and Raghu Ramakrishnan. Containment of conjunctive queries: Beyond relations as sets. *ACM Transactions on Database Systems*, 20(3):288–324, September 1995.

[JKV06] T. S. Jayram, Phokion G. Kolaitis, and Erik Vee. The containment problem for REAL conjunctive queries with inequalities. In Stijn Vansummeren, editor, *PODS*, pages 80–89. ACM, 2006.

[Kah01] Jeff Kahn. An entropy approach to the hard-core model on bipartite graphs. In *Combinatorics, Probability and Computing, Cambridge University Press*, volume 10. 2001.

[KST54] T. Kövári, V.T. Sós, and P. Turán. On a problem of k. zarankiewicz. *Colloq. Math.*, 3:50–57, 1954.

[Raz07] A. Razborov. Flag algebras. *J. Symbolic Logic*, 72(4):1239–01282, 2007.

[RS78] I. Z. Ruzsa and E. Szemerédi. Triple systems with no six points carrying three triangles. In *Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976), Vol. II*, volume 18 of *Colloq. Math. Soc. János Bolyai*, pages 939–945. North-Holland, Amsterdam, 1978.

[RV06] Benjamin Rossman and Erik Vee. Counting homomorphisms from trees and cycles. 2006. Unpublished manuscript.

[Vee06] Erik Vee, 2006. Personal communication.

[Yeu06] R. Yeung. *A First Course in Information Theory.* Springer, 2006.