

UNDECIDABILITY OF LINEAR INEQUALITIES IN GRAPH HOMOMORPHISM DENSITIES

HAMED HATAMI AND SERGUEI NORINE

1. INTRODUCTION

Many fundamental theorems in extremal graph theory can be expressed as algebraic inequalities between subgraph densities. As is explained below, for dense graphs, it is possible to replace subgraph densities with homomorphism densities. An easy observation shows that one can convert any algebraic inequality between homomorphism densities to a linear inequality. Inspired by the work of Freedman, Lovász and Schrijver [FLS07], in recent years a new line of research in the direction of treating and understanding these inequalities in a unified way has emerged. Razborov [Raz07] observed that a typical proof of an inequality in extremal graph theory between homomorphism densities of some fixed graphs involves only homomorphism densities of finitely many graphs. He states in [Raz08a] that in his opinion the most interesting general open question about asymptotic extremal combinatorics is whether every true linear inequality between homomorphism densities can be proved using a finite amount of manipulation with homomorphism densities of finitely many graphs. Although this question itself is not well defined, a natural precise refinement is whether the problem of determining the validity of a linear inequality between homomorphism densities is decidable. We show that it is not. Our result in particular answers various related questions by Razborov [Raz07], Lovász [Lov08], and Lovász and Szegedy [LS09].

An interesting recent result in extremal graph theory, proved in several different forms [FLS07, Raz07, LS09], says that every linear inequality between homomorphism densities follows from the positive semi-definiteness of a certain infinite matrix. As an immediate consequence, every algebraic inequality between the homomorphism densities follows from an infinite number of certain applications of the Cauchy-Schwarz inequality. This is consistent with the fact that many results in extremal graph theory are proved by one or more tricky applications of the Cauchy-Schwarz inequality. Lovász [Lov08] composed a collection of open problems in this area, and in Problem 17 he asks whether it is true or not that every algebraic inequality between homomorphism densities follows from a *finite* number of applications of this inequality. It is possible to rephrase this question in the language of quantum graphs defined by Freedman, Lovász and Schrijver [FLS07]. The validity

Received by the editors June 20, 2010 and, in revised form, August 29, 2010.

2010 *Mathematics Subject Classification*. Primary 05C25, 05C35, 12L05.

Key words and phrases. Graph homomorphism density, quantum graph, decidability, Artin's theorem.

The first author was supported in part by NSERC.

The second author was supported in part by NSF under Grant No. DMS-0701033.

©2010 American Mathematical Society
Reverts to public domain 28 years from publication

of a linear inequality between homomorphism densities corresponds to the positivity of a corresponding quantum graph. The square of a labeled quantum graph is trivially positive. In this language, Lovász’s question translates to the following statement: “Is it true that every positive quantum graph can be expressed as the sum of a finite number of squares of labeled quantum graphs?” The question in this form is stated by Lovász and Szegedy in [LS09]. In Theorem 2.4, we show that the answer is negative.

In [Raz07] Razborov introduced flag algebras, which provide a powerful formal calculus that captures many standard arguments in extremal combinatorics. He presented several questions about the linear inequalities between homomorphism densities among which is a question about a calculus introduced by him called the *Cauchy-Schwarz calculus*. This calculus, which allows trickier applications of the Cauchy-Schwarz inequality, can be used to prove the positivity of quantum graphs. He asks ([Raz07], Question 2) whether the Cauchy-Schwarz calculus is complete. We give a negative answer to this question by constructing positive quantum graphs whose positivity does not follow from this calculus.

A multivariate polynomial that takes only non-negative values over the reals is called *positive*. Our solutions to Lovász’s seventeenth problem and Razborov’s question about the Cauchy-Schwarz calculus are both based on reductions from real multivariate polynomials and they use the fact that there are positive polynomials that cannot be expressed as sums of squares of polynomials. Hence these answers are expected once one accepts the analogy to multivariate polynomials. However, Artin [Art27], solving Hilbert’s seventeenth problem, showed that every positive polynomial can be represented as a sum of squares of *rational functions*.

In Theorem 2.12 we prove that determining the validity of a linear inequality between homomorphism densities is undecidable. This reveals a major difference between the positivity of quantum graphs and the positivity of polynomials over reals as (for example by the celebrated work of Tarski [Tar48]) it is known that the latter is decidable. Furthermore we deduce from this theorem that the analogue of Artin’s solution to Hilbert’s seventeenth problem does not hold in the setting of quantum graphs. This in particular answers Problem 21 of Lovász’s list of open problems [Lov08].

Although our results show that not every algebraic inequality between homomorphism densities is a linear combination of a finite number of semi-definiteness inequalities, the positive semi-definite characterization is still a powerful approach for proving such inequalities. Razborov in [Raz08a] illustrated the power of this method by applying it to prove various results (some new and some known) in extremal combinatorics. Razborov [Raz08a] and Lovász and Szegedy [LS09] observed that it is possible to use this method to verify every linear inequality between homomorphism densities within an arbitrarily small error term. As this result suggests, the positive semi-definiteness method is extremely useful in proving bounds for problems in extremal combinatorics: Razborov [Raz08a] showed that a straightforward application of this method substantially improves the previously known bound for Turán’s function of K_4^3 , one of the most important problems in extremal combinatorics.

2. PRELIMINARIES

In this paper all graphs are simple and finite. For a graph G , let $V(G)$ and $E(G)$, respectively denote the set of the vertices and the edges of G . The unique graph with no vertices is denoted by \emptyset . The *density* of a graph H in a graph G , denoted by $t_{\text{inj}}(H; G)$, is the probability that a random *embedding* of the vertices of H in the vertices of G maps every edge of H to an edge of G . The *homomorphism density* of H in G , denoted by $t(H; G)$, is the probability that a random mapping (not necessarily injective) from the vertices of H to the vertices of G maps every edge of H to an edge of G . We define $t_{\text{inj}}(\emptyset; G) := t(\emptyset; G) := 1$, for every graph G .

Although $t(H; G)$ itself is an object of interest, extremal graph theory more often concerns $t_{\text{inj}}(H; G)$. However, the following simple lemma from [LS06] shows that these two quantities are close up to an error term of $o(1)$, and hence are equivalent asymptotically.

Lemma 2.1 ([LS06]). *For every two graphs H and G ,*

$$|t(H; G) - t_{\text{inj}}(H; G)| \leq \frac{1}{|V(G)|} \binom{|V(H)|}{2} = o_{|V(G)| \rightarrow \infty}(1).$$

Many important results in extremal graph theory can be expressed as algebraic inequalities between homomorphism densities. For example, Goodman’s theorem [Goo59], which generalizes the classical Mantel-Turán Theorem, says that for every graph G , $t(K_3; G) \geq 2t(K_2; G)^2 - t(K_2; G)$. Note that if $H_1 \dot{\cup} H_2$ denotes the disjoint union of two graphs H_1 and H_2 , we have $t(H_1 \dot{\cup} H_2; G) = t(H_1; G)t(H_2; G)$. This observation allows us to convert any algebraic inequality between homomorphism densities to a linear inequality. For example one can restate Goodman’s theorem as $t(K_3; G) - 2t(K_2 \dot{\cup} K_2; G) + t(K_2; G) \geq 0$.

A *partially labeled graph* is a graph in which some of the vertices are labeled by distinct natural numbers (there may be any number of unlabeled vertices). Let \mathcal{F} denote the set of all partially labeled graphs up to a label-preserving isomorphism. A partially labeled graph in which all vertices are labeled is called a *fully labeled graph*. The product of two partially labeled graphs H_1 and H_2 , denoted by $H_1 \cdot H_2$, is defined by taking their disjoint union, and then identifying vertices with the same label (if multiple edges arise, only one copy is kept). Clearly this multiplication is associative and commutative, and thus turns \mathcal{F} into a commutative semi-group. For every finite subset L of natural numbers, let \mathcal{F}_L denote the sub-semi-group of all partially labeled graphs whose set of labels is exactly L . Note that \mathcal{F}_\emptyset is the set of all finite graphs with no labels, and the product of two graphs in \mathcal{F}_\emptyset is their disjoint union.

We extend the definition of homomorphism density to partially labeled graphs in the following way. Consider a finite set $L \subset \mathbb{N}$, a partially labeled graph $H \in \mathcal{F}_L$, a graph G , and a map $\phi : L \rightarrow V(G)$. Then $t(H; G, \phi)$ is defined to be the probability that a random map from $V(H)$ to $V(G)$ is a homomorphism conditioned on the event that the labeled vertices are mapped according to ϕ . Note that for every two partially labeled graphs $H_1, H_2 \in \mathcal{F}_L$, a graph G , and a map $\phi : L \rightarrow V(G)$,

$$(2.2) \quad t(H_1 \cdot H_2; G, \phi) = t(H_1; G, \phi)t(H_2; G, \phi).$$

A *graph parameter* is a function that maps every graph to a real number. For example, given any graph G , the function $f : H \mapsto t(H; G)$ is a graph parameter. Freedman, Lovász, and Schrijver [FLS07] proved that it is possible to characterize

the graph parameters that are defined in a similar fashion using some positive semi-definiteness and rank conditions. Since then, various similar characterizations have been found. In particular, the following statement is proved in [LS09]: Let f be a graph parameter. There exists a sequence of graphs $\{G_n\}_{n \in \mathbb{N}}$ such that $\lim_{n \rightarrow \infty} t(H; G_n) = f(H)$, for every H , if and only if the following conditions are satisfied:

- (i) $f(\emptyset) = 1$, and $f(H_1) = f(H_2)$ if H_1 is obtained from H_2 by adding isolated vertices.
- (ii) The $\mathcal{F} \times \mathcal{F}$ matrix whose entry in the intersection of the row corresponding to H_1 and the column corresponding to H_2 is $f(H_1 \cdot H_2)$ is positive semi-definite. (Here the labels of $H_1 \cdot H_2$ are removed.)

Consider real numbers $\alpha_1, \dots, \alpha_k$ and graphs H_1, \dots, H_k . We wish to investigate whether an inequality of the form

$$(2.3) \quad \alpha_1 t(H_1; G) + \dots + \alpha_k t(H_k; G) \geq 0$$

holds for all graphs G . The positive semi-definiteness characterization shows that it suffices to verify the validity of $\alpha_1 f(H_1) + \dots + \alpha_k f(H_k) \geq 0$, for all graph parameters f satisfying Conditions (i) and (ii). Note that if the $\mathcal{F} \times \mathcal{F}$ matrix in Condition (ii) was finite, then there would exist an algorithm for solving this problem using semi-definite programming (see [Lov03] for a survey on this topic). However, since this matrix is of infinite dimensions, in practice one can only restrict to a finite submatrix of it and hope that $\alpha_1 f(H_1) + \dots + \alpha_k f(H_k) \geq 0$ is still valid if the weaker condition that this submatrix is positive semi-definite is required. The examples discussed in Section 1 show that this method is indeed very powerful and in many cases one succeeds in finding a proof for such inequalities. Lovász in [Lov08] asks (Problem 17) whether the validity of (2.3) always follows from the positive semi-definiteness of a finite submatrix of the $\mathcal{F} \times \mathcal{F}$ matrix in Condition (ii). As we will see in Section 2.1, it is possible to reformulate this question in the language of quantum graphs.

A *labeled quantum graph* is an element of the algebra $\mathbb{R}[\mathcal{F}]$, i.e. it is a formal linear combination of partially labeled graphs, and if $f = \sum_{i=1}^n \alpha_i H_i \in \mathbb{R}[\mathcal{F}]$ and $g = \sum_{i=1}^m \beta_i F_i \in \mathbb{R}[\mathcal{F}]$, then $f \cdot g = \sum_{i=1}^n \sum_{j=1}^m \alpha_i \beta_j H_i \cdot F_j$. For every finite set of positive integers L , $\mathbb{R}[\mathcal{F}_L]$ is a subalgebra of $\mathbb{R}[\mathcal{F}]$. The elements of $\mathbb{R}[\mathcal{F}_\emptyset]$ are called *quantum graphs*.

Consider a labeled quantum graph $f = \sum_{i=1}^k \alpha_i H_i \in \mathcal{F}$, and for every $i \in [k]$, let L_i be the set of all labels appearing on H_i . For a graph G and a map $\phi : \bigcup_{i=1}^k L_i \rightarrow V(G)$, define $t(f; G, \phi) := \sum_{i=1}^k \alpha_i t(H_i; G, \phi|_{L_i})$. Let \mathcal{K} be the ideal of $\mathbb{R}[\mathcal{F}]$ generated by elements of the form $F - \emptyset$, where F is a possibly labeled 1-vertex graph. Note that \mathcal{K} is the linear subspace of $\mathbb{R}[\mathcal{F}]$ spanned by elements of the form $F - H$, where F is obtained from H by adding a possibly labeled isolated vertex. Hence $t(f; G, \phi) = 0$ if $f \in \mathcal{K}$, and the function $t(\cdot; G, \phi)$ is a well-defined map from the quotient algebra $\mathcal{A} := \mathbb{R}[\mathcal{F}]/\mathcal{K}$ to \mathbb{R} .

Consider a finite set $L \subset \mathbb{N}$, a graph G , and a map $\phi : L \rightarrow V(G)$. It follows from (2.2) that $f \mapsto t(f; G, \phi)$ defines a homomorphism from $\mathbb{R}[\mathcal{F}_L]$ to \mathbb{R} , and hence it is also a well-defined homomorphism from $\mathcal{A}_L := \mathbb{R}[\mathcal{F}_L]/(\mathcal{K} \cap \mathbb{R}[\mathcal{F}_L])$ to \mathbb{R} .

For every finite set of positive integers L , let the linear map $[\cdot]_L : \mathbb{R}[\mathcal{F}] \rightarrow \mathbb{R}[\mathcal{F}]$ be defined by unlabeled the vertices whose labels are *not in* L . Note that this map is not an algebra homomorphism, as it does not respect the product. However

$[\cdot]_L$ maps \mathcal{K} to \mathcal{K} , so we can consider $[\cdot]_L$ as a linear map from \mathcal{A} to itself. We abbreviate $[\cdot]_\emptyset$ to $[\cdot]$.

2.1. Lovász’s seventeenth problem. We say that a labeled quantum graph $f \in \mathbb{R}[\mathcal{F}_L]$ is *positive* and write $f \geq 0$ if for every graph G and every $\phi : L \rightarrow V(G)$ we have $t(f; G, \phi) \geq 0$. By the discussion above, we can extend the definition of positivity to \mathcal{A}_L and further to the whole of \mathcal{A} .

Note that $g^2 \geq 0$, for every $g \in \mathbb{R}[\mathcal{F}_L]$. Furthermore for every subset $S \subseteq L$, the map $[\cdot]_S$ preserves positivity. It follows that $[[g^2]] \geq 0$, for every $g \in \mathbb{R}[\mathcal{F}]$. Hence one possible approach to prove an inequality of the form (2.3) is to express $f = \sum_{i=1}^k \alpha_i H_i$ as a sum of squares of labeled quantum graphs, i.e. to find labeled quantum graphs g_1, \dots, g_m such that $\sum_{i=1}^k \alpha_i H_i = \sum_{i=1}^m [[g_i^2]]$. Lovász’s seventeenth problem asks whether every positive quantum graph can be expressed in this form. In Section 4 we prove the following theorem which answers this question in the negative. We say that $x \in \mathcal{A}_\emptyset$ is *expressible as a sum of squares* if there exist $g_1, \dots, g_m \in \mathcal{A}$ such that $x = [[\sum_{i=1}^m g_i^2]]$.

Theorem 2.4. *There exists a positive quantum graph which cannot be expressed as a sum of squares.*

Whitney [Whi32, Theorem 5a] has shown that the functions $t_{\text{inj}}(H; \cdot)$, H connected, $V(H) > 1$, are algebraically independent. Equivalently, the functions $t(H; \cdot)$, H connected, $V(H) > 1$, are algebraically independent, as one can straightforwardly verify that a non-trivial algebraic relation between the functions $t_{\text{inj}}(H; \cdot)$ would imply a non-trivial algebraic relation between functions $t(H; \cdot)$. (In fact, it is shown in [ELS79] that the functions $t(H; \cdot)$ are independent in an even stronger sense.) It follows that

$$(2.5) \quad \mathcal{K} \cap \mathbb{R}[\mathcal{F}_\emptyset] = \{f \in \mathbb{R}[\mathcal{F}_\emptyset] : t(f; G) = 0 \text{ for every graph } G\}.$$

Therefore Theorem 2.4 is equivalent to the existence of a positive quantum graph x such that for every collection $g_1, \dots, g_m \in \mathbb{R}[\mathcal{F}]$, there exists a graph G so that

$$t(x; G) \neq t\left(\sum_{i=1}^m [[g_i^2]]; G\right).$$

2.2. Artin’s theorem. Note that Theorem 2.4 is reminiscent of Hilbert’s classical theorem that there exist positive multivariate real polynomials that cannot be expressed as sums of squares of polynomials. Hilbert in the seventeenth problem of his celebrated list of open problems asked “Given a multivariate polynomial that takes only non-negative values over the reals, can it be represented as a sum of squares of *rational functions*?” In 1927 Emil Artin [Art27] answered this question in the affirmative. Note that Artin’s theorem is equivalent to the fact that for every multivariate polynomial p that takes only non-negative values over the reals, there exist polynomials $q \neq 0$ and r , each expressible as a sum of squares of polynomials, such that $qp = r$. Our proof of Theorem 2.4 relies on the above-mentioned theorem of Hilbert. Hence it is very natural to wonder whether the analogue of Artin’s theorem holds for quantum graphs. Indeed Lemma 2.6 below (proved in Section 6) shows that the validity of such a statement would imply a simple finitary characterization of positive quantum graphs.

Lemma 2.6. *If $x \in \mathcal{A}_\emptyset$ satisfies $gx = h$ for some positive $g, h \in \mathcal{A}_\emptyset$ with $g \neq 0$, then x is positive.*

In Section 4 we prove the following theorem which shows that the analogue of Artin's theorem for quantum graphs does not hold.

Theorem 2.7. *There exists positive $x \in \mathcal{A}_\emptyset$ such that there are no $g, h \in \mathcal{A}_\emptyset$, each expressible as a sum of squares, with $g \neq 0$, so that $gx = h$.*

Note that Theorem 2.7 implies Theorem 2.4. However since our proof of Theorem 2.7 is based on the undecidability result proved in Theorem 2.12 below and hence does not provide any explicit examples, we give a separate constructive proof of Theorem 2.4.

Remark 2.8. In [Lov08], Problem 21, Lovász asked “Is it true that for every positive quantum graph x , there exist quantum graphs g and h , each expressible as a sum of squares of labeled quantum graphs, so that $x+gx = h$?” It follows from Theorem 2.7 that the answer to this question is also negative.

2.3. Razborov's Cauchy-Schwarz Calculus. Consider $f_1, f_2 \in \mathcal{A}_L$, and a subset $T \subseteq L$. Razborov ([Raz07], Theorem 3.14) has shown that

$$(2.9) \quad \llbracket f_1^2 \rrbracket_T \llbracket f_2^2 \rrbracket_T \geq \llbracket f_1 f_2 \rrbracket_T^2.$$

The *Cauchy-Schwarz calculus* is defined in [Raz07] in the language of flag algebras, but can be reformulated as follows.

Definition 2.10. The Cauchy-Schwarz calculus operates with statements of the form $f \geq 0$ with $f \in \mathcal{A}$ and has axioms

A1: $f^2 \geq 0$ for every f ;

A2: $\llbracket f_1^2 \rrbracket_T \llbracket f_2^2 \rrbracket_T - \llbracket f_1 f_2 \rrbracket_T^2 \geq 0$ for f_1, f_2 and T as in (2.9).

The inference rules of the Cauchy-Schwarz calculus are

R1:

$$\frac{f \geq 0 \quad g \geq 0}{\alpha f + \beta g \geq 0} \quad (\alpha, \beta \geq 0),$$

R2:

$$\frac{f \geq 0 \quad g \geq 0}{fg \geq 0},$$

R3:

$$\frac{f \geq 0}{\llbracket f \rrbracket_T \geq 0} \quad (T \subseteq \mathbb{N}),$$

for $f, g \in \mathcal{A}$. We say that $f \in \mathcal{A}$ is *CS-positive* if the statement $f \geq 0$ is provable in the Cauchy-Schwarz calculus.

The original definition of the Cauchy-Schwarz calculus in [Raz07] appears to differ from the one presented here. However, as we shall discuss in Appendix A, the two definitions are equivalent in that a statement can be proven in the original calculus if and only if it can be proven using the one stated in Definition 2.10.

Answering a question of Razborov, in Section 4 we show that the Cauchy-Schwarz calculus is not complete by proving the following theorem.

Theorem 2.11. *There exists a positive $f \in \mathcal{A}_\emptyset$ which is not CS-positive.*

2.4. A tenth problem: Undecidability. The proofs of both Theorems 2.4 and 2.11 rely on reductions from the setting of multivariate polynomials. It follows from Artin’s solution to Hilbert’s 17th problem (see Theorem 2.1.12 in [PD01]) that every positive $p \in \mathbb{Q}[x_1, \dots, x_n]$ can be expressed as a sum of the form $\sum_{i=1}^m a_i q_i^2$, where $a_i \in \mathbb{Q}_+$ and $q_i \in \mathbb{Q}(x_1, \dots, x_n)$ for every $1 \leq i \leq m$. Note that this in particular shows that the problem of determining whether a multivariate polynomial with rational coefficients is positive or not is decidable. Indeed, given such a polynomial, one can search for expressing it as a sum of the form $\sum_{i=1}^m a_i q_i^2$ (there are countably many of those sums), and in parallel check its non-negativity on rational points. Eventually either one will find a way to express the polynomial as a sum of squares of rational functions or a rational point on which the polynomial takes a negative value will be found. The decidability of positivity of a polynomial with rational coefficients follows also from the well-known work of Tarski [Tar48]. In the following theorem we show that the problem of determining the positivity of a quantum graph is undecidable.

Theorem 2.12. *The following problem is undecidable.*

- **INSTANCE:** *A positive integer k , finite graphs H_1, \dots, H_k , and integers a_1, \dots, a_k .*
- **QUESTION:** *Does the inequality $a_1 t(H_1; G) + \dots + a_k t(H_k; G) \geq 0$ hold for every graph G ?*

For two graphs H and G , let $\text{hom}(H; G)$ denote the number of homomorphisms from H to G . Note that $\text{hom}(H; G) = |V(G)|^{|V(H)|} t(H; G)$. In [KRar] it is observed that the undecidability of the following problem follows from the undecidability of a similar problem in database theory [IR95]. Given a positive integer k , finite graphs H_1, \dots, H_k , and integers a_1, \dots, a_k , it is undecidable whether the following inequality holds for all graphs G :

$$a_1 \text{hom}(H_1; G) + \dots + a_k \text{hom}(H_k; G) \geq 0.$$

Note that Theorem 2.12 in particular implies this result. Indeed, since $t(H; G) = \frac{\text{hom}(H; G)}{\text{hom}(K_1; G)^{|V(H)|}}$ for all graphs H and G , it is possible to express any linear inequality in homomorphism densities as an algebraic inequality in homomorphism numbers which in turn can be converted into a linear inequality.

3. SOME AUXILIARY FACTS

We start by defining some new notation and proving auxiliary facts. The *induced homomorphism density* of H in G , denoted by $t_{\text{ind}}(H; G)$, is the probability that a random map from the vertices of H to the vertices of G preserves both adjacency and non-adjacency. The two functions $t(H, \cdot)$ and $t_{\text{ind}}(H, \cdot)$ are related by

$$(3.1) \quad t(H; \cdot) = \sum_{\substack{F \supseteq H \\ V(F)=V(H)}} t_{\text{ind}}(F; \cdot)$$

and a Möbius inversion formula

$$(3.2) \quad t_{\text{ind}}(H; \cdot) = \sum_{\substack{F \supseteq H \\ V(F)=V(H)}} (-1)^{|E(F) \setminus E(H)|} t(F; \cdot).$$

For a partially labeled graph H , define the quantum graph

$$\text{ind}(H) := \sum_{\substack{F \supseteq H \\ V(F) = V(H)}} (-1)^{|E(F) \setminus E(H)|} F.$$

The labeled quantum graphs $\text{ind}(H)$ enjoy certain orthogonality properties. Indeed if the restriction of two partially labeled graphs $H_1, H_2 \in \mathcal{F}_L$ to the labeled vertices are different, then we have

$$(3.3) \quad \text{ind}(H_1) \cdot \text{ind}(H_2) = 0.$$

In Section 2 we defined the homomorphism density of a graph H in a graph G . Sometimes we shall work in a slightly more general setting that allows G to have a non-uniform distribution on its set of vertices. More precisely, let H be a partially labeled graph with the set of labels $L \subset \mathbb{N}$. Let G be another graph, \mathbf{y} be a probability measure on the vertices of G and $\phi : L \rightarrow V(G)$ be a map. Define the random mapping h from the vertices of H to the vertices of G by mapping every unlabeled vertex of H to a vertex of G independently and according to the probability measure \mathbf{y} , and mapping the labeled vertices according to ϕ . Then $t(H; G, \mathbf{y}, \phi)$ is the probability that h defines a homomorphism from H to G . Also $t_{\text{inj}}(H; G, \mathbf{y}, \phi)$ and $t_{\text{ind}}(H; G, \mathbf{y}, \phi)$ are defined similarly.

Remark 3.4. Consider a graph G and a probability measure \mathbf{y} on $V(G)$. For every positive integer n , construct the graph G_n in the following way. For every vertex v in $V(G)$, put $\lfloor \mathbf{y}(v)n \rfloor$ “copies” of v in G_n . There is an edge between two vertices in G_n if and only if they are copies of adjacent vertices in G . It is easy to see that for every graph H , we have $\lim_{n \rightarrow \infty} t(H; G_n) = t(H; G, \mathbf{y})$. Hence a quantum graph f is positive if and only if $t(f; G, \mathbf{y}) \geq 0$ for every graph G and every probability distribution \mathbf{y} on $V(G)$. Similarly it is easy to see that a labeled quantum graph f is positive if and only if $t(f; G, \mathbf{y}, \phi)$ is always non-negative.

Consider a graph H with $V(H) = [k]$. For every $j \in [k]$, let $H_j \in \mathcal{F}_{[k]}$ be obtained from H by adding an unlabeled clone v of the vertex j to this graph, i.e. v is adjacent exactly to the neighbors of the vertex j . Let H'_j be obtained from H_j by adding the edge $\{v, j\}$. Then $\varphi_H : \mathbb{R}[x_1, \dots, x_k] \rightarrow \mathbb{R}[\mathcal{F}_{[k]}]$ is the unique algebra homomorphism defined by $\varphi_H(x_j) := \text{ind}(H_j) + \text{ind}(H'_j)$, for every $j \in [k]$.

Consider a graph G and a probability measure \mathbf{y} on $V(G)$. Let S denote the set of all maps $h : V(H) \rightarrow V(G)$ that preserve both adjacency and non-adjacency. Consider a map $\phi : V(H) \rightarrow V(G)$. If $\phi \notin S$, then $t(\varphi_H(x_j); G, \mathbf{y}, \phi) = 0$, for every $j \in [k]$. Hence in this case $t(\varphi_H(p); G, \mathbf{y}, \phi) = 0$, for every polynomial $p \in \mathbb{R}[x_1, \dots, x_k]$.

Next consider a map $\phi \in S$. Then for every $j \in [k]$, $t(\varphi_H(x_j); G, \mathbf{y}, \phi)$ is the probability that a random (according to \mathbf{y}) extension ψ of ϕ in H_j preserves the adjacencies and non-adjacencies except maybe between j and its clone. For every $j \in [k]$, let $\alpha_j(\phi)$ be the probability that the map obtained from ϕ by replacing $\phi(j)$ by a random vertex in G chosen according to the probability measure \mathbf{y} belongs to S . Since the unlabeled vertex of H_j is a clone of the vertex j , we have $t(\varphi_H(x_j); G, \mathbf{y}, \phi) = \alpha_j(\phi)$.

We conclude that for every map $\phi : [k] \rightarrow V(G)$, and every polynomial $p \in \mathbb{R}[x_1, \dots, x_k]$,

$$(3.5) \quad t(\varphi_H(p); G, \mathbf{y}, \phi) = \begin{cases} p(\alpha_1(\phi), \dots, \alpha_k(\phi)) & \phi \in S, \\ 0 & \phi \notin S. \end{cases}$$

The next lemma follows immediately from (3.5).

Lemma 3.6. *Given a graph H with $V(H) = [k]$ and a positive polynomial p in k variables, the labeled quantum graph $\varphi_H(p)$ is positive.*

Let P_k denote the set of all positive *homogeneous* polynomials in k variables. Let Σ_k denote the set of those homogeneous polynomials in k variables that can be expressed as sums of squares of polynomials with real coefficients. Clearly $\Sigma_k \subseteq P_k$. Let $\Delta_k = P_k \setminus \Sigma_k$. Hilbert [Hil88] showed that for $k \geq 3$, Δ_k is not empty.

An $x_0 \in \mathbb{R}^k$ is called [Rez00] a *bad point* for polynomial $p \in P_k$ if x_0 is a root of every polynomial q such that $q^2 p \in \Sigma_k$.

Lemma 3.7. *For every $k \geq 4$, there exists an even $p \in \Delta_k$ such that $x_1^2 x_2^2 \dots x_k^2$ divides p and $(1, 0, \dots, 0)$ is a bad point for p .*

Proof. Consider the polynomial $p := x_1^2 x_2^2 \dots x_k^2 S(x_2, x_3, x_4)$, where $S(x, y, z)$ is an even homogeneous polynomial in Δ_3 , e.g. $S(x, y, z) := x^4 y^2 + y^4 z^2 + z^4 x^2 - 3x^2 y^2 z^2$. A short argument that this particular polynomial belongs to Δ_3 is provided for example in [Lam05, p. 519].

Note that p is trivially positive. Suppose for the contradiction that $q^2 p = \sum_{i=1}^m q_i^2$ for polynomials q_i while $q(1, 0, \dots, 0) = c \neq 0$ for a polynomial q of degree d . Note that $q(1, 0, \dots, 0) = c$ implies that the monomial x_1^d appears with coefficient c in q . Hence the component of x_1^{2d+2} in $q^2 p$ is $c^2 (x_2^2 \dots x_k^2) S(x_2, x_3, x_4)$. Denoting by \bar{q}_i the component of x_1^{d+1} in q_i , we must have $c^2 (x_2^2 \dots x_k^2) S(x_2, x_3, x_4) = \sum_{i=1}^m \bar{q}_i^2$. Note that as the left side of the equality is divisible by x_j for each $2 \leq j \leq k$ we can readily see (by setting $x_j = 0$) that every \bar{q}_i is divisible by every x_j and consequently by $x_2 \dots x_k$. It now follows that $S(x_2, x_3, x_4) \in \Sigma_3$, which is a contradiction. \square

4. PROOFS OF THEOREMS 2.4 AND 2.11

For finite $L \subset \mathbb{N}$, $f \in \mathbb{R}[\mathcal{F}_L]$ and a graph G on n vertices we say that f is *G-sos* if for every $\phi : L \rightarrow V(G)$ there exist polynomials $p_1, \dots, p_m \in \mathbb{R}[y_1, \dots, y_n]$ with *non-negative coefficients*, and $q_1, \dots, q_m \in \mathbb{R}[y_1, \dots, y_n]$ such that for all probability distributions \mathbf{y} ,

$$t(f; G, \mathbf{y}, \phi) = \sum_{i=1}^m p_i(\mathbf{y}) q_i(\mathbf{y})^2.$$

The definition of *G-sos* extends to \mathcal{A} . The following lemma, which is a key step in the proofs of Theorems 2.4 and 2.11, relates the squares of labeled quantum graphs to the squares of polynomials.

Lemma 4.1. *Let G be a graph. Then*

- (i) f^2 is *G-sos* for every $f \in \mathcal{A}$;
- (ii) $\llbracket f^2 \rrbracket_T \llbracket g^2 \rrbracket_T - \llbracket fg \rrbracket_T^2$ is *G-sos* for all finite $T \subseteq \mathbb{N}$ and $f, g \in \mathcal{A}$;
- (iii) if f and g are *G-sos*, then $\alpha f + \beta g$ is *G-sos* for all $\alpha, \beta \in \mathbb{R}_+$;

- (iv) if f and g are G -sos, then fg is G -sos;
- (v) if f is G -sos, then $\llbracket f \rrbracket_T$ is G -sos for all finite $T \subseteq \mathbb{N}$.

Proof. To verify assertion (i), take $m = 1$, $p_1 = 1$, and $q_1 = t(f; G, \mathbf{y}, \phi)$ in the definition of a G -sos element. For (ii), without loss of generality, we assume that $f, g \in \mathcal{A}_L$ for some finite $L \supseteq T$. Denoting $A := t(\llbracket f^2 \rrbracket_T \llbracket g^2 \rrbracket_T - \llbracket fg \rrbracket_T^2; G, \mathbf{y}, \phi)$, we have

$$\begin{aligned} A &= \left(\sum_{\psi} \left(\prod_{v \in L \setminus T} y_{\psi(v)} \right) t^2(f; G, \mathbf{y}, \psi) \right) \left(\sum_{\psi} \left(\prod_{v \in L \setminus T} y_{\psi(v)} \right) t^2(g; G, \mathbf{y}, \psi) \right) \\ &\quad - \left(\sum_{\psi} \left(\prod_{v \in L \setminus T} y_{\psi(v)} \right) t(f; G, \mathbf{y}, \psi) t(g; G, \mathbf{y}, \psi) \right)^2 \\ &= \frac{1}{2} \sum_{\psi_1} \sum_{\psi_2} \left(\prod_{v \in L \setminus T} y_{\psi_1(v)} \right) \left(\prod_{v \in L \setminus T} y_{\psi_2(v)} \right) \\ &\quad \times (t(f; G, \mathbf{y}, \psi_1) t(g; G, \mathbf{y}, \psi_2) - t(f; G, \mathbf{y}, \psi_2) t(g; G, \mathbf{y}, \psi_1))^2, \end{aligned}$$

where the summations over ψ, ψ_1 and ψ_2 are over all maps from L to $V(G)$, which coincide with ϕ on T . Assertions (iii) and (iv) are trivial. Finally, for (v) if $f \in \mathbb{R}[\mathcal{F}_L]$, then we have

$$t(\llbracket f \rrbracket_T; G, \mathbf{y}, \phi) = \sum_{\psi} \left(\prod_{v \in L \setminus T} y_{\psi(v)} \right) t(f, \psi; G, \mathbf{y}),$$

where the summation is over all maps ψ from L to $V(G)$ which coincide with ϕ on T . □

Consider a graph H . A set $W \subseteq V(H)$ is called *homogeneous* in H if for every two distinct vertices $u, v \in W$, $N(u) \setminus W \neq N(v) \setminus W$, where $N(u)$ and $N(v)$ respectively denote the set of the neighbors of u and v . We call a graph H *stringent* if it does not contain any homogeneous subsets W with $1 < |W| \leq |V(H)| - 1$, and furthermore does not allow any non-trivial automorphisms. Note that if H is stringent, then in particular the identity map is the only map from H to itself that preserves both adjacency and non-adjacency.

Stringent graphs serve as the foundation for all our constructions, and so we will need the following simple lemma.

Lemma 4.2. *For every $k \geq 6$, there exists a stringent graph on k vertices.*

Proof. Let the graph H on $k \geq 6$ vertices be obtained from a triangle with vertices $v_1 v_2 v_3$ by adding a path $v_3 v_4 \dots v_k$ and finally by joining v_k to v_2 and v_3 . See Figure 1.

Suppose first that $W \subseteq V(H)$ with $1 < |W| \leq |V(H)| - 1$ is homogeneous. Then $|W| \neq |V(H)| - 1$, as no vertex of H has $|V(H)| - 1$ neighbors. The graph H is 2-connected, and therefore W has at least two neighbors outside of W . It follows that W must lie in the intersection of the neighborhoods of two vertices in $V(H) \setminus W$. This allows for a single possibility, namely $W = \{v_2, v_3\}$, which is clearly non-homogeneous. Similarly, one can routinely verify that H allows no non-trivial automorphisms, and thus is stringent. □

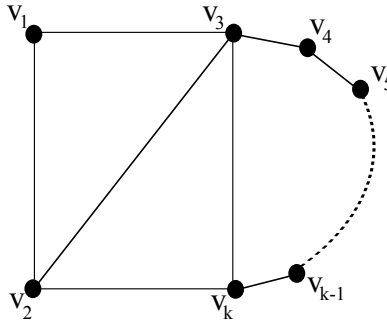


FIGURE 1. A stringent graph on $k \geq 6$ vertices.

Proof of Theorem 2.4. Let H be a stringent graph on $k \geq 4$ vertices. Let $V(H) = [k]$. By (3.5) for every homogeneous polynomial $p \in \mathbb{R}[x_1, \dots, x_k]$, and every map $\phi : [k] \rightarrow V(H)$,

$$(4.3) \quad t(\varphi_H(p); H, \mathbf{y}, \phi) = \begin{cases} p(y_1, \dots, y_k) & \phi = \text{id}, \\ 0 & \phi \neq \text{id}. \end{cases}$$

Since the probability that a random map ϕ picked according to the probability measure \mathbf{y} is equal to the identity map is $y_1 y_2 \dots y_k$, it follows from (4.3) that

$$(4.4) \quad t(\llbracket \varphi_H(p) \rrbracket; H, \mathbf{y}) = y_1 y_2 \dots y_k p(y_1, \dots, y_k).$$

Let p be a homogeneous polynomial in k variables such that the even homogeneous polynomial $x_1^2 x_2^2 \dots x_k^2 p(x_1^2, \dots, x_k^2)$ satisfies the assertion of Lemma 3.7, e.g.

$$p(x_1^2, \dots, x_k^2) = x_2^4 x_3^2 + x_3^4 x_4^2 + x_4^4 x_2^2 - 3x_2^2 x_3^2 x_4^2.$$

By Lemma 3.6, the quantum graph $x := \llbracket \varphi_H(p) \rrbracket$ is positive. We claim that x satisfies the assertion of Theorem 2.4. Assume to the contrary that there exist labeled quantum graphs g_1, \dots, g_m , such that for every graph G ,

$$t(x; G) = \sum_{i=1}^m t(\llbracket g_i^2 \rrbracket; G).$$

Then by Remark 3.4, for every graph G and every probability distribution \mathbf{y} on the vertices of G ,

$$t(x; G, \mathbf{y}) = \sum_{i=1}^m t(\llbracket g_i^2 \rrbracket; G, \mathbf{y}).$$

By Lemma 4.1 there exist polynomials $p_1, \dots, p_m \in \mathbb{R}[y_1, \dots, y_k]$ with non-negative coefficients, and $q_1, \dots, q_m \in \mathbb{R}[y_1, \dots, y_k]$ so that

$$(4.5) \quad t(x; H, \mathbf{y}) = \sum_{i=1}^m p_i(\mathbf{y}) q_i(\mathbf{y})^2.$$

Hence by (4.4),

$$(4.6) \quad y_1 y_2 \dots y_k p(y_1, \dots, y_k) = \sum_{i=1}^m p_i(\mathbf{y}) q_i(\mathbf{y})^2.$$

Now consider arbitrary real numbers x_1, \dots, x_k not all of them zero. Set

$$(4.7) \quad \mathbf{y} := \left(\frac{x_1^2}{x_1^2 + \dots + x_k^2}, \dots, \frac{x_k^2}{x_1^2 + \dots + x_k^2} \right).$$

As p is homogeneous, we have

$$y_1 y_2 \dots y_k p(\mathbf{y}) = (x_1^2 + \dots + x_k^2)^{-\deg(p)-k} x_1^2 \dots x_k^2 p(x_1^2, x_2^2, \dots, x_k^2).$$

Substituting this in (4.6) and multiplying both sides by a large enough power of $(x_1^2 + \dots + x_k^2)$ shows that

$$(x_1^2 + x_2^2 + \dots + x_k^2)^N (x_1^2 x_2^2 \dots x_k^2 p(x_1^2, x_2^2, \dots, x_k^2)) \in \Sigma_k,$$

for some positive integer N . But this contradicts the assumption that $(1, 0, \dots, 0)$ is a bad point for $x_1^2 x_2^2 \dots x_k^2 p(x_1^2, \dots, x_k^2)$. \square

Proof of Theorem 2.11. The proof parallels the proof of Theorem 2.4. Let H, p and x be as in the proof of Theorem 2.4. Then x is positive. If x is CS-positive, Lemma 4.1 shows that $t(x; H, \mathbf{y}) = \sum_{i=1}^m p_i(\mathbf{y}) q_i(\mathbf{y})^2$ for polynomials $p_1, \dots, p_m \in \mathbb{R}[y_1, \dots, y_k]$ with non-negative coefficients, and $q_1, \dots, q_m \in \mathbb{R}[y_1, \dots, y_k]$. The rest of the proof proceeds as in that of Theorem 2.4. \square

5. UNDECIDABILITY

As we discussed in Section 2, determining the positivity of a polynomial with rational coefficients is decidable. However it follows from Matiyasevich’s solution to Hilbert’s tenth problem [Mat70] that if one restricts to integer-valued variables, this problem becomes undecidable. More precisely, given a multivariate polynomial with integer coefficients, the problem of determining whether it is non-negative for every assignment of integers (equivalently, positive integers) to its variables is undecidable. To prove the undecidability in Theorem 2.12 we will need the following simple consequence of this fact.

Lemma 5.1. *The following problem is undecidable.*

- INSTANCE: *A positive integer $k \geq 6$, and a polynomial $p(x_1, \dots, x_k)$ with integer coefficients.*
- QUESTION: *Do there exist $x_1, \dots, x_k \in \{1 - 1/n : n \in \mathbb{N}\}$ such that $p(x_1, \dots, x_k) < 0$?*

Proof. Consider a polynomial $q(y_1, \dots, y_k)$ with integer coefficients. Note that $q(y_1, \dots, y_k) \geq 0$ for all $y_1, \dots, y_k \in \mathbb{N}$ if and only if the polynomial with integer coefficients

$$p(x_1, \dots, x_k) := \left(\prod_{i=1}^k (1 - x_i)^{\deg(q)} \right) q \left(\frac{1}{1 - x_1}, \dots, \frac{1}{1 - x_k} \right)$$

is non-negative for all $x_1, \dots, x_k \in \{1 - 1/n : n \in \mathbb{N}\}$. Hence the problem is undecidable. \square

It is not a priori clear how the non-negativity of formulas of the form $a_1 t(H_1; \cdot) + \dots + a_k t(H_k; \cdot)$ is related to the non-negativity of a polynomial on *integers*. Indeed, if one considers a single graph H , then the set of all possible values of $t(H; \cdot)$ is everywhere dense in the interval $[0, 1]$. The key ingredient in the proof of Theorem 2.12 is the observation that there are relations between $t(H; \cdot)$ for different graphs H which are satisfied for an infinite, but sparse set of possible values of the

corresponding homomorphism densities. This can already be seen in the case of the relation between the edge homomorphism density and the triangle homomorphism density. Let $g(x) := 2x^2 - x$. As is mentioned in Section 1, Goodman [Goo59] proved that $t(K_3; G) \geq g(t(K_2; G))$, for every graph G . Bollobás improved Goodman’s bound to the following.

Theorem 5.2 (Bollobás [Bol76]). *For every graph G , and every positive integer t , if*

$$t(K_2; G) \in \left[1 - \frac{1}{t}, 1 - \frac{1}{t+1}\right],$$

then we have $t(K_3; G) \geq L(t(K_2; G))$, where

$$(5.3) \quad L(x) := \frac{3t^2 - t - 2}{t(t+1)}x - \frac{2(t-1)}{t+1}.$$

Note that for every positive integer t , on the interval $\left[1 - \frac{1}{t}, 1 - \frac{1}{t+1}\right]$, L is the linear function that coincides with g on the endpoints. Razborov [Raz08b] has recently proven the exact lower bound for $t(K_3; G)$ in terms of $t(K_2; G)$, but Bollobás’s result suffices for our purpose. Let $L : [0, 1) \rightarrow \mathbb{R}$ be the continuous piecewise linear function defined on $\left[1 - \frac{1}{t}, 1 - \frac{1}{t+1}\right]$ by (5.3) for every positive integer t . By Theorem 5.2, for every graph G , $(t(K_2; G), t(K_3; G)) \in R$ where $R \subset [0, 1]^2$ is the region defined as

$$R := \{(x, y) \in [0, 1]^2 : y \geq L(x)\}.$$

The examples of complete graphs show that Goodman’s bound is tight when

$$t(K_2; \cdot) \in \left\{1 - \frac{1}{n} : n \in \mathbb{N}\right\},$$

and on the other hand, Theorem 5.2 shows that it is *not* tight on the rest of the interval $[0, 1)$. Hence the algebraic expression $t(K_3; G) - g(t(K_2; G))$ can be equal to 0 if and only if $t(K_2; \cdot) \in \left\{1 - \frac{1}{n} : n \in \mathbb{N}\right\}$. This already reveals the connection to Lemma 5.1 and suggests a direction for proving Theorem 2.12.

Lemma 5.4. *Let p be a polynomial in variables x_1, \dots, x_k . Let M be the sum of the absolute values of the coefficients of p multiplied by $100 \deg(p)$. Define $q \in \mathbb{R}[x_1, \dots, x_k, y_1, \dots, y_k]$ as*

$$q := p \prod_{i=1}^k (1 - x_i)^6 + M \left(\sum_{i=1}^k y_i - g(x_i) \right).$$

Then the following are equivalent:

- (a): $q(x_1, \dots, x_k, y_1, \dots, y_k) < 0$ for some $x_1, \dots, x_k, y_1, \dots, y_k$ with $(x_i, y_i) \in R$ for every $1 \leq i \leq k$;
- (b): $p(x_1, \dots, x_k) < 0$ for some $x_1, \dots, x_k \in \left\{1 - \frac{1}{n} : n \in \mathbb{N}\right\}$.

Proof. If (b) holds, then for each x_i we have $(x_i, g(x_i)) \in R$ and setting $y_i := g(x_i)$ gives $q(x_1, \dots, x_k, y_1, \dots, y_k) < 0$. Therefore (b) implies (a).

Suppose now that (a) holds. Decreasing y_i decreases the value of q and thus we assume without loss of generality that $y_i = L(x_i)$. Let

$$\begin{aligned} \tilde{q}(x_1, x_2, \dots, x_k) &= q(x_1, \dots, x_k, y_1, \dots, y_k) \\ &= p(x_1, \dots, x_k) \prod_{i=1}^k (1 - x_i)^6 + M \left(\sum_{i=1}^k L(x_i) - g(x_i) \right) < 0. \end{aligned}$$

For every $1 \leq i \leq k$, let t_i be a positive integer such that $x_i \in \left[1 - \frac{1}{t_i}, 1 - \frac{1}{t_i+1}\right]$. Fixing t_1, \dots, t_k we assume that x_1, \dots, x_k are chosen in the corresponding intervals to minimize \tilde{q} . We claim that in this case $x_i \in \{1 - 1/n : n \in \mathbb{N}\}$ for every $1 \leq i \leq k$. Suppose for a contradiction that $x_i \in \left(1 - \frac{1}{t_i}, 1 - \frac{1}{t_i+1}\right)$, for some i . By the choice of x_i we have $\frac{\partial \tilde{q}}{\partial x_i} = 0$. Hence since

$$\left| \frac{\partial}{\partial x_i} \left(p(x_1, \dots, x_k) \prod_{i=1}^k (1 - x_i)^6 \right) \right| \leq 7 \frac{M}{100} (1 - x_i)^5 \leq \frac{M}{12t_i^5},$$

we must have

$$(5.5) \quad \frac{1}{12t_i^5} \geq |L'(x_i) - g'(x_i)| = \left| \frac{3t_i^2 - t_i - 2}{t_i(t_i + 1)} - 4x_i + 1 \right| = 4 \left| \frac{t_i^2 - 1/2}{t_i(t_i + 1)} - x_i \right|.$$

Let $z = \frac{t_i^2 - 1/2}{t_i(t_i + 1)}$. We can rewrite (5.5) as $|z - x_i| \leq \frac{1}{48t_i^5}$. Note that $L'(z) = g'(z)$, $L'(x) - g'(x)$ is monotone on the interval between x_i and z , and $L(z) - g(z) = 1/(2t_i^2(t_i + 1)^2)$. It follows that

$$\begin{aligned} L(x_i) - g(x_i) &\geq (L(z) - g(z)) - |L'(x_i) - g'(x_i)||z - x_i| \\ &\geq \frac{1}{2t_i^2(t_i + 1)^2} - \frac{1}{48t_i^{10}} \geq \frac{1}{8t_i^4} - \frac{1}{48t_i^4} \geq \frac{1}{10t_i^4}. \end{aligned}$$

Finally we have

$$\tilde{q}(x_1, x_2, \dots, x_k) \geq -\frac{M}{100} (1 - x_i)^6 + M(L(x_i) - g(x_i)) \geq M \left(\frac{1}{10t_i^4} - \frac{1}{100t_i^6} \right) \geq 0,$$

which is a contradiction. Therefore the claim that $x_i \in \{1 - 1/n : n \in \mathbb{N}\}$ for every $1 \leq i \leq k$ holds. In this case we have

$$0 > \tilde{q}(x_1, x_2, \dots, x_k) = p(x_1, x_2, \dots, x_k) \prod_{i=1}^k (1 - x_i)^6,$$

which shows that (b) holds. □

Define the map

$$\tau : \mathbb{R}[x_1, \dots, x_k, y_1, \dots, y_k] \rightarrow \mathbb{R}[v_1, \dots, v_k, e_1, \dots, e_k, t_1, \dots, t_k]$$

in the following way. For every polynomial q , let $\tau(q)$ be obtained from q by substituting for every i , e_i/v_i^2 and t_i/v_i^3 instead of x_i and y_i , respectively, and multiplying the resulting rational function by $\prod_{i=1}^k v_i^{3 \deg(q)}$ so that it becomes a polynomial.

Given a graph H with $V(H) = [k]$, we define the labeled quantum graphs $V_i, E_i, T_i \in \mathbb{R}[\mathcal{F}_{[k]}]$ in the following way. For every positive integer m and every $j \in [k]$, let $H_{j,m} \in \mathcal{F}_{[k]}$ be the graph on $k + m$ vertices obtained from H by adding a clique of size m and connecting each one of the vertices of this clique to

the neighbors of the vertex j . Then $V_j := \sum \text{ind}(H_{j,1} \cup F)$, $E_j := \sum_F \text{ind}(H_{j,2} \cup F)$, and $T_j := \sum_F \text{ind}(H_{j,3} \cup F)$, where all these sums are over different ways of joining the unlabeled vertices to the vertex j .

Let $\psi_H : \mathbb{R}[v_1, \dots, v_k, e_1, \dots, e_k, t_1, \dots, t_k] \rightarrow \mathbb{R}[\mathcal{F}_{[k]}]$ be the unique algebra homomorphism that satisfies $\psi_H(v_i) = V_i$, $\psi_H(e_i) = E_i$, and $\psi_H(t_i) = T_i$ for every $i \in [k]$.

Consider a graph G , and let S denote the set of all maps $h : V(H) \rightarrow V(G)$ that preserve both adjacency and non-adjacency. Consider a map $\phi : V(H) \rightarrow V(G)$. If $\phi \notin S$, then $t(\text{ind}(H); G, \phi) = 0$, which in particular shows that $t(V_i; G, \phi) = t(E_i; G, \phi) = t(T_i; G, \phi) = 0$ for every $i \in [k]$. Hence in this case $t(\psi_H(\tau(q)); G, \phi) = 0$ for every polynomial $q \in \mathbb{R}[x_1, \dots, x_k, y_1, \dots, y_k]$.

Next consider a map $\phi \in S$. For every $j \in [k]$, let $U_j(\phi)$ be the subgraph of G induced on the set of vertices v for which the map obtained from ϕ by replacing $\phi(j)$ by v belongs to S . Note that for every $j \in [k]$,

$$t(V_j; G, \phi) = \frac{|U_j|}{|V(G)|}, \quad t(E_j; G, \phi) = t(K_2; U_j) \frac{|U_j|^2}{|V(G)|^2},$$

$$t(T_j; G, \phi) = t(K_3; U_j) \frac{|U_j|^3}{|V(G)|^3}.$$

This shows that

$$\frac{t(E_j; G, \phi)}{t(V_j; G, \phi)^2} = t(K_2; U_j) \quad \text{and} \quad \frac{t(T_j; G, \phi)}{t(V_j; G, \phi)^3} = t(K_3; U_j).$$

Recalling the definitions of τ and ψ_H , we conclude that

$$(5.6) \quad t(\psi_H(\tau(q)); G, \phi) = q(t(K_2; U_1), \dots, t(K_2; U_k), t(K_3; U_1), \dots, t(K_3; U_k)) \prod_{j=1}^k \left(\frac{|U_j|}{|V(G)|} \right)^{3 \deg(q)}$$

for every polynomial $q \in \mathbb{R}[x_1, \dots, x_k, y_1, \dots, y_k]$.

Claim 5.7. Let $q \in \mathbb{R}[x_1, \dots, x_k, y_1, \dots, y_k]$ be such that $q(x_1, \dots, x_k, y_1, \dots, y_k) \geq 0$ for all $x_1, \dots, x_k, y_1, \dots, y_k$ with $(x_i, y_i) \in R$ for every $1 \leq i \leq k$. Then $\psi_H(\tau(q))$ is a positive labeled quantum graph for every graph H with $V(H) = [k]$.

Proof. Theorem 5.2 implies that $(t(K_2; U_j), t(K_3; U_j)) \in R$, for every $j \in [k]$. Now the claim follows from (5.6). \square

Claim 5.8. Let $q \in \mathbb{R}[x_1, \dots, x_k, y_1, \dots, y_k]$ be such that $q(x_1, \dots, x_k, y_1, \dots, y_k) < 0$ for some $x_i \in \{1 - 1/n : n \in \mathbb{N}\}$ and $y_i = g(x_i)$ for $1 \leq i \leq k$. Let H be a stringent graph with $V(H) = [k]$. Then there exists a graph G such that

$$t(\llbracket \psi_H(\tau(q)) \rrbracket; G) < 0.$$

Proof. Let $n_1, \dots, n_k \in \mathbb{N}$ be chosen so that q becomes negative by setting $x_i := 1 - 1/n_i$ and $y_i := 2x_i^2 - x_i$ for all $i \in [k]$. Define G to be the graph obtained from H by replacing the vertex j of H by a clique of size n_j for every $j \in [k]$. Let W_j be the set of the vertices of the clique that replaces the vertex j of H in G .

Let S denote the set of all maps $h : V(H) \rightarrow V(G)$ that preserve both adjacency and non-adjacency. Consider a map $\phi \in S$. It follows from the structure of G that for every j , $\{i : \phi(i) \in W_j\}$ is a homogeneous set in H , and since H is stringent it is of size at most 1. (Trivially, it cannot be all of $V(H)$). Hence for every j ,

$\{i : \phi(i) \in W_j\}$ is of size exactly 1. Since H is stringent, the identity map is the only isomorphism from H to itself. It follows that $\phi(j) \in W_j$, for every $j \in [k]$. Then for every $j \in [k]$, $U_j(\phi)$ is the restriction of G to W_j , which by definition of G is a clique of size n_j . Thus

$$t(K_2; U_j) = 1 - 1/n_j \quad \text{and} \quad t(K_3; U_j) = g(1 - 1/n_j),$$

which by (5.6) shows that

$$t(\psi_H(\tau(q)); G, \phi) = q(x_1, \dots, x_k, y_1, \dots, y_k) \prod_{j=1}^k \left(\frac{|n_j|}{|V(G)|} \right)^{3 \deg(q)} < 0.$$

Moreover if $\phi \notin S$, then $t(\psi_H(\tau(q)); G, \phi) = 0$. We conclude that $t(\llbracket \psi_H(\tau(q)) \rrbracket; G) < 0$. \square

Proof of Theorem 2.12. Consider an instance of the undecidable problem stated in Lemma 5.1, namely a polynomial p in variables x_1, \dots, x_k with integer coefficients. Construct the polynomial q in variables $x_1, \dots, x_k, y_1, \dots, y_k$ as in Lemma 5.4. Then Lemma 5.4 shows that $p(x_1, \dots, x_k) < 0$ for some $x_1, \dots, x_k \in \{1 - 1/n : n \in \mathbb{N}\}$ if and only if $q(x_1, \dots, x_k, y_1, \dots, y_k) < 0$ for some $x_1, \dots, x_k, y_1, \dots, y_k$ with $(x_i, y_i) \in R$ for every $1 \leq i \leq k$. By Claims 5.7 and 5.8, determining the latter is equivalent to determining the validity of $t(\llbracket \psi_H(\tau(q)) \rrbracket; \cdot) \geq 0$, where H is a stringent graph on k vertices. Such graphs exist and can be explicitly constructed by Lemma 4.2. \square

6. PROOF OF THEOREM 2.7

Before giving the proofs of Lemma 2.6 and Theorem 2.7 we need to recall some facts about graphons. Note that if A_G is the adjacency matrix of a graph G , then for every graph H ,

$$(6.1) \quad t(H; G) = \mathbb{E} \prod_{uv \in E(H)} A(x_u, x_v),$$

where $\{x_u : u \in V(H)\}$ are independent uniform random variables taking values in $\{1, 2, \dots, |V(G)|\}$. Let \mathcal{W}_0 denote the set of bounded symmetric measurable functions of the form $w : [0, 1]^2 \rightarrow [0, 1]$. The elements of \mathcal{W}_0 are called *graphons*. For every graph H , and every graphon $w \in \mathcal{W}_0$, define by analogy with (6.1),

$$(6.2) \quad t(H; w) := \int \prod_{uv \in E(H)} w(x_u, x_v) \prod_{v \in V(H)} dx_v.$$

This definition can be extended linearly to define $t(x; w)$, for every quantum graph x . For every graph G , we define a graphon $w_G \in \mathcal{W}_0$ as follows. Without loss of generality assume that $V(G) = [n]$. Then $w_G(x, y) := A_G(\lceil xn \rceil; \lceil yn \rceil)$ if $x, y \in (0; 1]$, and if $x = 0$ or $y = 0$, then $w_G(x, y) := 0$. By (6.1) and (6.2), for every quantum graph x and graph G , we have $t(x; G) = t(x; w_G)$.

A graph sequence $\{G_i\}_{i \in \mathbb{N}}$ is called convergent if for every graph H , the limit $\lim_{i \rightarrow \infty} t(H; G_i)$ exists. It is shown in [LS06] that for every convergent graph sequence $\{G_i\}_{i \in \mathbb{N}}$, there exists a graphon w such that $\lim_{i \rightarrow \infty} t(H; G_i) = t(H; w)$, for every graph H . On the other hand for every graphon w , it is easy to construct a convergent graph sequence $\{G_i\}_{i \in \mathbb{N}}$ such that $t(H; w) = \lim_{i \rightarrow \infty} t(H; G_i)$, for every graph H .

Proof of Lemma 2.6. If x is not positive, then there exists a graphon w such that $t(x; w) < 0$. Since $g \neq 0$, by (2.5) there exists a graphon w' satisfying $t(g; w') \neq 0$. Now note that by (6.2), $t(g; \alpha w' + (1 - \alpha)w)$ is a polynomial in α . This polynomial is not identically 0 as it is not equal to zero on $\alpha = 1$. Hence there are arbitrarily small $\alpha > 0$ for which $t(g; \alpha w' + (1 - \alpha)w) \neq 0$. By taking a sufficiently small such α , we obtain a graphon $w'' := \alpha w' + (1 - \alpha)w$ that satisfies both

$$t(g; w'') > 0 \quad \text{and} \quad t(x; w'') < 0.$$

This in particular implies that $t(gx; w'') < 0$, contradicting $gx = h$ and the assumption that h is expressible as a sum of squares. \square

Proof of Theorem 2.7. Suppose to the contrary that for every positive $x \in \mathcal{A}$, there exist $g, h \in \mathcal{A}$, $g \neq 0$, each expressible as sums of squares, such that $gx = h$. We will show this would imply that given a quantum graph f with rational coefficients, the problem of determining the validity of $f \geq 0$ is decidable, contradicting Theorem 2.12.

Given a collection of partially labeled graphs $\mathcal{G} = (G_1, \dots, G_k)$, an integer m and a matrix $A = (a_{ij} : i \in [m], j \in [k])$ define

$$z(\mathcal{G}, m, A) := \left[\sum_{i=1}^m \left(\sum_{j=1}^k a_{ij} G_j \right)^2 \right].$$

This quantum graph can be expressed as a linear combination of graphs of the form $\llbracket G_{j_1} \cdot G_{j_2} \rrbracket$, where $j_1, j_2 \in [k]$, with the coefficients polynomial in the entries of A . Note that connected non-isomorphic graphs are algebraically independent as elements of \mathcal{A}_\emptyset , and every graph as an element of \mathcal{A}_\emptyset is equal to the product of its connected components. It follows that for a fixed quantum graph f with rational coefficients, a fixed collection of partially labeled graphs \mathcal{G} and an integer m , the system

$$(6.3) \quad z(\mathcal{G}, m, A)f = z(\mathcal{G}, m, B) \quad \text{and} \quad z(\mathcal{G}, m, A) \neq 0$$

can be expressed as a (computable) system of polynomial equations and inequalities with rational coefficients on the entries of A and B . Therefore, it is possible to decide whether there exist matrices A and B with real entries solving this system.

Hence in order to decide the validity of $f \geq 0$, one enumerates finite graphs G and checks the validity of $t(f; G) \geq 0$ on each graph. In parallel, one enumerates all pairs (\mathcal{G}, m) , where \mathcal{G} is a finite sequence of finite partially labeled graphs and m is an integer, and for each such pair checks whether there exists a solution to (6.3). \square

APPENDIX A. ORIGINAL FORMULATION
OF RAZBOROV'S CAUCHY-SCHWARZ CALCULUS

In the formulation given in [Raz07], the Cauchy-Schwarz calculus contains an additional axiom and an additional inference rule. The axiom can be stated here as:

- $\text{ind}(H) \geq 0$ for every partially labeled graph H .

This axiom can be derived in the Cauchy-Schwarz calculus presented in Definition 2.10 as follows. Let H' be a graph obtained from H by assigning new labels to the previously unlabeled vertices of H , so that all vertices of H' are labeled. Then

we have $(\text{ind}(H'))^2 = \text{ind}(H')$. It follows that $\text{ind}(H) = \llbracket (\text{ind}(H'))^2 \rrbracket_T$, where T is the set of the labels used on the vertices of H . Thus the above axiom follows from the axiom A1 and the inference rule R3 in Definition 2.10.

Let f be a labeled quantum graph, and let H be a graph, with all vertices of H labeled. We say that f is H -rooted if there exists a positive integer k , partially labeled graphs G_1, G_2, \dots, G_k and real numbers $\alpha_1, \dots, \alpha_k$ such that $f = \sum_{i=1}^k \alpha_i \text{ind}(G_i)$ and in every G_i the subgraph induced by the labeled vertices is equal to H . An additional inference rule from [Raz07] can now be stated as follows:

- Let f be an H -rooted labeled quantum graph with $f \geq 0$. Let H' be a graph with all vertices labeled, such that H is an induced subgraph of H' as a labeled graph. Then $f \cdot \text{ind}(H') \geq 0$.

Note that this inference rule is subsumed in inference rule R3 of Definition 2.10.

Let us further note that in [Raz07] the product of two labeled quantum graphs is only defined if both of the graphs are H -rooted for some H . In our framework two partially labeled graphs correspond to the same element of \mathcal{A} if one is obtained from another by adding isolated, possibly labeled, vertices. Therefore we can consider every labeled quantum graph as a linear combination of partially labeled graphs, all of which have exactly the same set of labeled vertices. Thus every labeled quantum graph f can be written in a form $f = \sum_H f_H$, where the summation is taken over all labeled graphs H with $V(H) = [l]$ for some positive integer l , so that the vertices of H are labeled in the natural way, and each f_H is H -rooted. One can routinely deduce from the definitions that $f \geq 0$ if and only if $f_H \geq 0$ for every H . Further, it follows from (3.3) that if $f = \sum_H f_H$ and $g = \sum_H g_H$ are as above, then $fg = \sum_H f_H g_H$. Consequently, our multiplication inference rule could be restricted as in [Raz07] to multiplying only H -rooted quantum graphs.

ACKNOWLEDGEMENTS

The authors wish to thank Alexander Razborov for many enlightening discussions. They also wish to thank Swastik Kopparty, László Lovász and the anonymous referee for their valuable comments.

REFERENCES

- [Art27] Emil Artin, *Über die Zerlegung definiter Funktionen in Quadrate*, Abh. Math. Sem. Hamburg **5** (1927), 100–115.
- [Bol76] Béla Bollobás, *Relations between sets of complete subgraphs*, Proceedings of the Fifth British Combinatorial Conference (Univ. Aberdeen, Aberdeen, 1975), Congressus Numerantium, No. XV, Utilitas Math., Winnipeg, Man., 1976, pp. 79–84. MR0396327 (53:195)
- [ELS79] Paul Erdős, László Lovász, and Joel Spencer, *Strong independence of graphcopy functions*, Graph theory and related topics (Proc. Conf., Univ. Waterloo, Waterloo, Ont., 1977), Academic Press, New York, 1979, pp. 165–172. MR538044 (81b:05060)
- [FLS07] Michael Freedman, László Lovász, and Alexander Schrijver, *Reflection positivity, rank connectivity, and homomorphism of graphs*, J. Amer. Math. Soc. **20** (2007), no. 1, 37–51 (electronic). MR2257396 (2007h:05153)
- [Goo59] A. W. Goodman, *On sets of acquaintances and strangers at any party*, Amer. Math. Monthly **66** (1959), 778–783. MR0107610 (21:6335)
- [Hil88] David Hilbert, *Ueber Büschel von binären Formen mit vorgeschriebener Functional-determinante*, Math. Ann. **33** (1888), no. 2, 227–236. MR1510539
- [IR95] Yannis E. Ioannidis and Raghuram Ramakrishnan, *Containment of conjunctive queries: beyond relations as sets*, ACM Trans. Database Syst. **20** (1995), no. 3, 288–324.

- [KRar] Swastik Kopparty and Benjamin Rossman, *The homomorphism domination exponent*, European J. Combin. (to appear).
- [Lam05] T. Y. Lam, *Introduction to quadratic forms over fields*, Graduate Studies in Mathematics, vol. 67, American Mathematical Society, Providence, RI, 2005. MR2104929 (2005h:11075)
- [Lov03] László Lovász, *Semidefinite programs and combinatorial optimization*, Recent advances in algorithms and combinatorics, CMS Books Math./Ouvrages Math. SMC, vol. 11, Springer, New York, 2003, pp. 137–194. MR1952986 (2003k:90046)
- [Lov08] ———, *Graph homomorphisms: Open problems*, manuscript (2008).
- [LS06] László Lovász and Balázs Szegedy, *Limits of dense graph sequences*, J. Combin. Theory Ser. B **96** (2006), no. 6, 933–957. MR2274085 (2007m:05132)
- [LS09] ———, *Random graphons and a weak positivstellensatz for graphs*, arXiv.org:0902.1327 (2009).
- [Mat70] Ju. V. Matijasevič, *The Diophantineness of enumerable sets*, Dokl. Akad. Nauk SSSR **191** (1970), 279–282. MR0258744 (41:3390)
- [PD01] Alexander Prestel and Charles N. Delzell, *Positive polynomials*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 2001, From Hilbert’s 17th problem to real algebra. MR1829790 (2002k:13044)
- [Raz07] Alexander A. Razborov, *Flag algebras*, J. Symbolic Logic **72** (2007), no. 4, 1239–1282. MR2371204 (2008j:03040)
- [Raz08a] ———, *On 3-hypergraphs with forbidden 4-vertex configurations*, manuscript (2008).
- [Raz08b] ———, *On the minimal density of triangles in graphs*, Combin. Probab. Comput. **17** (2008), no. 4, 603–618. MR2433944 (2009i:05118)
- [Rez00] Bruce Reznick, *Some concrete aspects of Hilbert’s 17th Problem*, Real algebraic geometry and ordered structures (Baton Rouge, LA, 1996), Contemp. Math., vol. 253, Amer. Math. Soc., Providence, RI, 2000, pp. 251–272. MR1747589 (2001i:11042)
- [Tar48] Alfred Tarski, *A Decision Method for Elementary Algebra and Geometry*, RAND Corporation, Santa Monica, Calif., 1948. MR0028796 (10,499f)
- [Whi32] Hassler Whitney, *The coloring of graphs*, Ann. of Math. (2) **33** (1932), no. 4, 688–718. MR1503085

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY, PRINCETON, NEW JERSEY 08544
Current address: School of Computer Science, McGill University, Montréal, Quebec, Canada
E-mail address: hatami@cs.mcgill.ca

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY, PRINCETON, NEW JERSEY 08544
E-mail address: snorin@math.princeton.edu