Available online at www.sciencedirect.com
ScienceDirect
Journal of Combinatorial Theory, Series B 98 (2008) 146–163
Journal of Combinatorial Theory Series B
www.elsevier.com/locate/jctb

Generalized quasirandom graphs

László Lovász [a], Vera T. Sós [b,1]

[a] *Microsoft Research, Redmond, WA, USA*
[b] *Rényi Institute, Budapest, Hungary*

Received 11 January 2006
Available online 25 October 2007

Abstract

We prove that if a sequence of graphs has (asymptotically) the same distribution of small subgraphs as a generalized random graph modeled on a fixed weighted graph H, then these graphs have a structure that is asymptotically the same as the structure of H. Furthermore, it suffices to require this for a finite number of subgraphs, whose number and size is bounded by a function of $|V(H)|$.

Keywords: Quasirandom graph; Homomorphism; Graph algebra; Convergent graph sequence

1. Introduction

Quasirandom (also called pseudorandom) graphs were introduced by Thomason [10] and Chung, Graham and Wilson [3]. These graphs have many properties that true random graphs have.

To be more precise, a sequence $(G_n\colon\ n = 1, 2, \dots)$ of graphs is called *quasirandom with density* p (where $0 < 1 < p$), if for every simple finite graph F, the number of copies of F in G_n is asymptotically $|V(G_n)|^{|V(F)|}p^{|E(F)|}$ (this is the asymptotic number of copies of F in a random graph with edge probability p; we consider labeled copies, so for example the number of copies of K_2 in G_n in $2|E(G_n)|$).

It turns out that this definition implies many other properties that are familiar from the theory of random graphs; for example, almost all degrees are about pn, almost all codegrees are about p^2n, all cuts with $\Theta(n)$ nodes on both sides have edge-density about p etc. Many of these

E-mail address: lovasz@cs.elte.hu (L. Lovász).
[1] Research of the author was supported in part by OTKA grants T032236, T038210, T042750.

0095-8956/$ – see front matter
doi:10.1016/j.jctb.2007.06.005

characterize quasirandom graphs, and this fact provides many equivalent ways to define a quasirandom sequence [3,8–10]. Quasirandomness is closely related to Szemerédi's lemma [7]. One of the most surprising facts proved in [3] is that it is enough to require the condition about the number of copies of F for just two graphs, namely $F = K_2$ and C_4.

Consider a weighted graph H on q nodes, with a weight $\alpha_i > 0$ associated with each node and a weight $0 \leqslant \beta_{ij} \leqslant 1$ associated with each edge ij. We may assume that H is complete with a loop at every node, since the missing edges can be added with weight 0. A *generalized random graph* $\mathbf{G}(n; H)$ with model H is generated as follows. We take $[n] = \{1, \dots, n\}$ as its node set. We partition $[n]$ into q sets $V_1, \dots, V_q$, by putting node u in V_i with probability α_i, and connecting each pair $u \in V_i$ and $v \in V_j$ with probability β_{ij} (all these decisions are made independently). A *generalized quasirandom graph sequence* (G_n) with model H is defined by the property that for every fixed finite graph F, the number of copies of F in G_n is asymptotically the same as the number of copies of F in a generalized random graph $\mathbf{G}(N, H)$ on $N = |V(G_n)|$ nodes.

One can define, more generally, *convergent* sequences of graphs (G_n) by the property that for every fixed finite graph F, the number of copies of F in G_n, appropriately normalized, is convergent [1,2], and a limit object can be assigned to every convergent sequence [6]. Generalized quasirandom sequences are convergent sequences with the special property that their limit can be expressed as a finite weighted graph. This motivates the following two basic questions concerning generalized quasirandom graphs:

(a) Is it enough to require the condition concerning the number of copies of F for a finite set of graphs F_i (depending on H)?
(b) Is the structure of a generalized quasirandom graph similar to a generalized random graph in the following sense? Its nodes can be partitioned into q classes $V_1, \dots, V_q$ of sizes $\alpha_1 N, \dots, \alpha_q N$ so that the graph spanned by V_i is quasirandom with density $\beta_{i,i,}$ and the bipartite graph formed by the edges between V_i and V_j is quasirandom with density β_{ij} (for the modification of the definition of quasirandomness to bipartite graphs, see the next section).

In this paper we answer both questions in the affirmative. The main tool is to formulate the conditions in terms of homomorphisms of graphs, and then invoke the tool of graph algebras borrowed from a recent paper of Freedman, Lovász and Schrijver [4].

Recent results about limits of graph sequences [6] and distances of graphs [2] (see [1] for a survey) yield another proof of (b), and in fact in a more general form characterizing convergent graph sequences as convergence in a suitable metric. However, this proof does not seem to imply the affirmative answer to (a), i.e., the finiteness of the number of test graphs needed.

One may expect that the following converse to (a) also has an affirmative answer: "Let $F_1, \dots, F_k$ be finite graphs and $0 \leqslant a_1, \dots, a_k \leqslant 1$. Assume that every sequence (G_n) of graphs for which the density of F_i in G_n converges to a_i for every i is convergent. Does it follow that every such sequence (G_n) is generalized quasirandom?" Recently B. Szegedy and Lovász (unpublished) found a counterexample. The question of how far the notion of quasirandomness can be relaxed so that (a) remains true is open.

Quasirandom graph sequences have several other characterizations, in terms of cuts, eigenvalues, Szemerédi partitions, etc. Most of these extend to H-quasirandom graph sequences, and even to the more general setting of convergent graph sequences: several results that guarantee (b) under various "multiway cut" conditions are proved in [2]. (The most notable exception is the spectrum, which does not carry enough information to determine the structure of the graph as in (b).) It would be interesting to find analogues of (a) for these other characterizations.

2. Preliminaries and results

2.1. Homomorphisms and quasirandom graphs

For any simple unweighted graph F and weighted graph H, we define

$$\hom(F,H)=\sum_{\psi:V(F)\to V(H)}\alpha_\psi\beta_\psi$$

where

$$\alpha_\psi=\prod_{i\in V(F)}\alpha_{\psi(i)}$$

and

$$\beta_\psi=\prod_{ij\in E(F)}\beta_{\psi(i)\psi(j)}.$$

We consider an unweighted graph as a weighted graph where all nodeweights and edgeweights are 1, and if there is no edge, the weight is 0. In this case, $\hom(F,H)$ counts the number of homomorphisms of F into H (adjacency-preserving maps of $V(F)$ into $V(H)$).

A sequence (G_n) of simple unweighted graphs is *quasirandom with density* p, if for every simple graph F

$$\frac{\hom(F,G_n)}{|V(G_n)|^{|V(F)|}}\to p^{|E(F)|}\quad(n\to\infty).$$

If, for every $n\geqslant 1$, G_n is a (ordinary) random graph $\mathbf{G}(n,p)$, then the sequence (G_n) is quasirandom with probability 1.

It will be convenient to think of a bipartite graph H as having an "upper" bipartition class $U(H)$ and a "lower" bipartition class $W(H)$. For two simple, unweighted bipartite graphs F and H, let $\hom'(F,H)$ denote the number of those homomorphisms of F into H that map $U(F)$ to $U(H)$ and $W(F)$ to $W(H)$. A sequence (G_n) of bipartite graphs is *bipartite quasirandom with density* p, if for every simple bipartite graph F

$$\frac{\hom(F,G_n)}{|U(G_n)|^{|U(F)|}|W(G_n)|^{|W(F)|}}\to p^{|E(F)|}\quad(n\to\infty).$$

The following result from [3] will be important for us:

Theorem 2.1. *A sequence (G_n) of graphs is quasirandom with density p if and only if*

$$\frac{\hom(K_2,G_n)}{|V(G_n)|^2}\to p\quad(n\to\infty)$$

and

$$\frac{\hom(C_4,G_n)}{|V(G_n)|^4}\to p^4\quad(n\to\infty).$$

An analogous result holds for bipartite quasirandom graphs.

2.2. Generalized quasirandom graphs

Let $G_1, G_2, \dots$ be unweighted graphs and H, a weighted graph on $V(H) = [q]$ such that $\sum_{i \in V(H)} \alpha_i = 1$ and $0 \leqslant \beta_{ij} \leqslant 1$ for every $i, j \in V(H)$. We may assume that H is complete (with loops at each node), since the missing edges can be added with weight 0. We say that the sequence (G_n) is *H-quasirandom,* if for every unweighted, simple graph F,

$$\frac{\hom(F, G_n)}{|V(G_n)|^{|V(F)|}} \to \hom(F, H). \tag{1}$$

In the special case when H is a single node, with a loop with weight p, we get the definition of a quasirandom sequence.

One way to construct a H-quasirandom sequence is the following. Take n nodes (where n is very large), and partition them into q classes $V_1, \dots, V_q$ (where $|V(H)| = \{1, \dots, q\}$) so that

$$|V_i| \approx \alpha_i n.$$

For every i, insert on the nodes of V_i a quasirandom graph with density β_{ii}, and for every $i \neq j$, insert between the nodes of V_i and V_j a bipartite quasirandom graph with density β_{ij}.

Our main result is that the converse is true:

Theorem 2.2. *Let H be a weighted graph with $V(H) = [q]$, node weights $(\alpha_i\colon i = 1, \dots, q)$ and edge weights $(\beta_{ij}\colon i, j = 1, \dots, q)$. Let $(G_n, n = 1, 2, \dots)$ be a H-quasirandom sequence of unweighted simple graphs. Then for every n there exists a partition $V(G_n) = \{V_1, \dots, V_q\}$ such that*

(a) $\frac{|V_i|}{|V(G_n)|} \to \alpha_i$ $(i = 1, \dots, q)$,
(b) *the subgraph of G_n induced by V_i is a quasirandom graph sequence with edge density β_{ii} for all $i = 1, \dots, q$, and*
(c) *the bipartite subgraph between V_i and V_j is a quasirandom bipartite graph sequence with edge-density β_{ij} for all $i, j = 1, \dots, q$, $i \neq j$.*

It is not hard to see that conversely, every graph sequence (G_n) with structure (a)–(b)–(c) is H-quasirandom. The proof of Theorem 2.2 will also show the following fact, which can be thought of as a generalization of Theorem 2.1:

Theorem 2.3. *A weighted graph H on q nodes is H-quasirandom if and only if*

$$\frac{\hom(F, G_n)}{|V(G_n)|^{|V(F)|}} \to \hom(F, H)$$

for every simple graph F with at most $q + (10q)^q$ nodes.

The bound on the size of the graphs F can certainly be improved, but to determine the exact minimum seems very difficult. The main point is that it depends only on the number of nodes in H, not on the edgeweights or nodeweights.

2.3. Plan of the proof

Suppose that we have a (small) weighted model graph H with $V(H) = [q]$ and a (huge) simple graph G_n with $V(G_n) = [n]$. We would like to classify the nodes of G_n, so that each class corresponds to a node of H. Given a node u of G_n, we would like to find a corresponding node i of H.

A first idea is to look at the degree $d_{G_n}(u)$ of u, and match it with a node i of corresponding degree; the degree of i should be defined as $d_H(i)=\sum_j \alpha_j\beta_{ij}$ (where the β_{ij} are the edgeweights in H), and we want that $d_{G_n}(u)\approx d_H(i)n$.

It is not too hard to show that for "most" nodes of G_n there is a node in H for which this degree condition holds (with an error tending to 0 as $n\to\infty$). Consider the star S_m with m nodes, then

$$\hom(S_m,H)=\sum_{i=1}^{q}\alpha_i d_H(i)^{m-1},\qquad \frac{\hom(S_m,G_n)}{n^m}=\frac{1}{n}\sum_{u=1}^{n}\left(\frac{d_{G_n}(u)}{n}\right)^{m-1}.$$

From the fact that these two exponential functions of m are close for every m, it follows that the bases for the exponentials can be matched up: about $\alpha_i n$ terms on the right-hand side must be close to $d_H(i)$, for $i=1,\dots,q$.

The trouble is that H may have several nodes with the same degree. To refine our argument, we look at larger neighborhoods; in other words, we count not only the number of edges incident with u, but also the number of triangles hanging from u, the number of paths of length 2 starting at u etc.

In general, let F be any (simple, unweighted) graph with $V(F)=[k]$, where node 1 is considered as a special "root." We count the number of homomorphisms of F into G_n that map 1 onto u, to get a number $\hom_u(F,G_n)$. The corresponding quantity for a weighted graph H is

$$\hom_i(F,H)=\sum_{\substack{\psi:V(F)\to[q]\\ \psi(1)=i}}\prod_{m=2}^{k}\alpha_{\psi(m)}\prod_{jm\in E(F)}\beta_{\psi(j)\psi(m)}$$

for $i\in V(H)$. (We take those terms in the definition of $\hom(F,H)$ with $\psi(1)=i$ and omit the factor α_i. Multiplying this number by n^{q-1}, we get asymptotically $\hom_v(F,\mathbf{G}(n,H))$ for any $v\in V_i$.) Note that

$$\sum_{i\in[q]}\alpha_i\hom_i(F,H)=\hom(F,H),$$

and

$$\sum_{u\in[n]}\hom_u(F,G_n)=\hom(F,G_n).$$

We want to match u with a node i of H for which $\hom_u(F,G_n)\approx\hom_i(F,H)n^{q-1}$ for all F.

Consider the vectors

$$h_F=\bigl(\hom_1(F,H),\dots,\hom_q(F,H)\bigr)\in\mathbb{R}^q.$$

There are infinitely many of these, but they live in a finite dimensional space $\mathbb{R}^q$. Suppose that $\{h_{F_1},\dots,h_{F_q}\}$ form a basis of $\mathbb{R}^q$, then we can express the vector $e_1=(1,0,\dots,0)$ as a linear combination of them:

$$e_1=\sum_{i=1}^{q}\lambda_i h_{F_i}.$$

Now consider the analogous vectors

$$g_F = \big(\hom_1(F, G_n)/n^{q-1}, \dots, \hom_n(F, G_n)n^{q-1}\big) \in \mathbb{R}^n,$$

and the linear combination

$$s = \sum_{i=1}^{q} \lambda_i g_{F_i}.$$

If a node u is "similar" to node 1 of H, then s_u should be about 1; if u is similar to some other node of H, then s_u should be close to 0. So the large entries of s should tell us which nodes of G_n should be matched with 1. We could find the nodes to be matched with $2, 3, \dots, q$ similarly.

To develop this idea to a proof, there are several difficulties. To show that for most nodes u of G_n, the sequence $(g_F(u))$ is similar to a sequence $(h_F(i))$ we have to extend our argument above. A convenient tool for this will be the language of *quantum graphs* and *graph algebras*, developed in [4–6].

The most substantial difficulty in filling out the details is the following. We assumed above that the vectors h_F span the whole space $\mathbb{R}^q$. This is not so in general; the trouble is caused by two (related but different) symmetries H may have: *twin nodes* and *automorphisms.* Of these, twins are easy to eliminate (see Section 3.4), but automorphisms cause a conceptual problem. For example, the model graph H may have a node-transitive automorphism group; then there is no way to distinguish between its nodes, and our whole scheme for finding a "match" for u fails.

The way out will be to use not one special node in F but q of them; if we fix a bijective map of these nodes onto $V(H)$, then this breaks any symmetry between the nodes of H. We will have to pay for this trick with a lot of technical details.

Let us mention one further difficulty, less serious but still nontrivial. Let F be a finite graph with multiple edges, and let F' be the simple graph obtained from F by forgetting about the edge multiplicities. Then $\hom(F', G_n) = \hom(F, G_n)$ (since the G_n are unweighted), but $\hom(F', H) \neq \hom(F, H)$ in general. So the sequence $\hom(F, G_n)/|V(G_n)|^{|V(F)|}$ is convergent, but its limit is $\hom(F', H)$ rather than $\hom(F, H)$. We started with using only simple graphs, but when we glue them together along more than one node, we may create multiple edges. In Section 3.5 we describe a construction from [6] that can be used to eliminate these.

3. Graph algebras

3.1. Quantum graphs

We introduce some formalism. A *quantum graph* is a formal finite linear combination (with real coefficients) of graphs. Quantum graphs form an (infinite dimensional) linear space $\mathcal{G}_0$. We can introduce a multiplication in this space: for two ordinary graphs, the product is defined as disjoint union; we extend this linearly to quantum graphs. This turns $\mathcal{G}_0$ into a commutative and associative algebra.

We extend these constructions to a slightly more complex situation. Fix a positive integer k. A *k-labeled graph* is a finite graph in which some of the nodes are labeled by numbers $1, \dots, k$ (a node can have at most one label). Two k-labeled graphs are *isomorphic*, if there is a label-preserving isomorphism between them. We denote by K_k the k-labeled complete graph with k nodes, and by E_k, the k-labeled graph with k nodes and no edges. $\emptyset$-labeled graphs are just ordinary graphs.

A *k-labeled quantum graph* is a formal finite linear combination (with real coefficients) of k-labeled graphs. Let $\mathcal{G}_k$ denote the (infinite dimensional) vector space of all k-labeled quantum graphs.

Let F_1 and F_2 be two k-labeled graphs. Their *product* F_1F_2 is defined as follows: we take their disjoint union, and then identify nodes with the same label. (Note that F_1F_2 may have multiple edges even F_1 and F_2 are simple.) Clearly this multiplication is associative and commutative. Extending this multiplication to k-labeled quantum graphs linearly, we get an associative and commutative algebra $\mathcal{G}_k$. The graph E_k with k labeled nodes and no edges is a unit element in $\mathcal{G}_k$.

3.2. Partial homomorphism functions

For every k-labeled graph F, weighted graph H, and $\varphi : [k] \to [q]$, we define

$$\hom_\varphi(F, H) = \sum_{\substack{\psi: V(F)\to[q]\\ \psi \text{ extends } \varphi}} \prod_{i\in V(F)\setminus[k]} \alpha_{\psi(i)} \prod_{ij\in E(F)} \beta_{\psi(i)\psi(j)}.$$

We extend the definition of $\hom_\varphi(x, H)$ to all $x \in \mathcal{G}_k$ linearly. If we fix a map $\varphi : [k] \to [q]$, then the map $\hom_\varphi(., H)$ will be multiplicative on $\mathcal{G}_k$. If F is a k-labeled graph, we also write $\hom_{i_1\dots i_k}$ instead of $\hom_\varphi$ where $\varphi(1) = i_1, \dots, \varphi(k) = i_k$.

Clearly $\hom_\varphi(F, H) \leqslant 1$ for every $\varphi : [k] \to V(H)$. So if $x = \sum_i \lambda_i F_i \in \mathcal{G}_k$, then

$$\left|\hom_\varphi(x, H)\right| = \left|\sum_i \lambda_i \hom_\varphi(F_i, H)\right| \leqslant \sum_i |\lambda_i| = N(x). \tag{2}$$

If G is an unweighted graph with n nodes, then the same argument gives that

$$\frac{|\hom_\varphi(x, G)|}{n^k} \leqslant N(x). \tag{3}$$

What will be important for us is that the right-hand side is independent of G.

3.3. Graph homomorphisms and algebra homomorphisms

Fix a weighted "model graph" H with $V(H) = [q]$, with nodeweights $\alpha_1, \dots, \alpha_q$ and edgeweights β_{ij}. The algebras $\mathcal{G}_k$ are independent of the model graph H, but we use the $\hom(., H)$ function to introduce additional structure.

First, for $k = 0$, we can define $\hom(x, H)$ for every quantum graph x, by extending it linearly from the generators. Then we have, for $x, y \in \mathcal{G}_0$,

$$\hom(x + y, H) = \hom(x, H) + \hom(y, H)$$

and

$$\hom(xy, H) = \hom(x, H)\hom(y, H),$$

so $\hom(x, H)$ is an algebra homomorphism from $\mathcal{G}_0$ into the reals.

The function $\hom(., H)$ is not multiplicative on $\mathcal{G}_k$ for $k \geqslant 1$, but for every fixed mapping $\phi : [k] \to V(H)$, the mapping $\hom_\phi(., H)$ is multiplicative. If we view $\mathbb{R}^{[q]^k}$ as an algebra (the direct product of q^k copies of $\mathbb{R}$), then we get an algebra homomorphism Ξ_k from $\mathcal{G}_k$ into $\mathbb{R}^{[q]^k}$. We denote by $\mathcal{N}_k$ the kernel of Ξ_k.

We can also use the hom(.,.) to introduce a bilinear form on $\mathcal{G}_k$ by

$$\langle x, y\rangle = \hom(xy, H).$$

In particular, we have

$$\langle F_1, F_2\rangle = \hom(F_1F_2, H)$$

for two ordinary graphs F_1 and F_2. It is not hard to see [4] that this bilinear form is semidefinite: $\langle x, x\rangle \geqslant 0$ for all x. So we can define

$$\|x\| = \langle x, x\rangle^{1/2}.$$

This value is a seminorm, but not a norm, because there will be quantum graphs x with $\|x\| = 0$. We write $x = y \pmod H$ if $\|x - y\| = 0$. It is not hard to show that this is equivalent to saying that $\langle x - y, z\rangle = 0$ for every $z \in \mathcal{G}_k$. A further equivalent formulation is that $\hom_\phi(x - y, H) = 0$ for every $\phi : [k] \to [q]$, i.e., $x - y \in \mathcal{N}_k$.

We can factor out $\mathcal{N}_k$, to obtain an algebra $\mathcal{G}_k/H = \mathcal{G}_k/\mathcal{N}_k$. The bilinear form $\langle .,.\rangle$ gives a positive definite inner product on $\mathcal{G}_k/H$. It was shown in [4] that this algebra is finite dimensional (see Corollary 3.2 below).

3.4. *Twins and automorphisms*

Let us think of $\mathbb{R}^{q^k}$ as vectors indexed by maps $\varphi : [k] \to [q]$. For every $x \in \mathcal{G}_k$, the vector $(\hom_\varphi(x, H)\colon\ \varphi \in [q]^k)$ is in this space. Can every vector in $\mathbb{R}^{q^k}$ be realized by some quantum graph x? The answer is "generically" in the affirmative, but not always. There are two (similar, but slightly different) reasons this.

We call two nodes $i, j \in [q]$ *twins*, if for every node $k \in [q]$, $\beta_{ik} = \beta_{jk}$ (note: the condition includes $k = i$ and $k = j$; the node weights α_i play no role in this definition).

Suppose that H is not twin-free, so that it has two twin nodes i and j. Then for any $x \in \mathcal{G}_1$, the numbers $\hom_i(x, H)$ and $\hom_j(x, H)$ differ by the same scalar, so not every vector in $\mathbb{R}^q$ can be realized.

This trouble is, however, easily eliminated. If H is not twin-free, we can identify the equivalence classes of twin nodes, define the node-weight α of a new node as the sum of the node-weights of its pre-images, and define the weight of an edge as the weight of any of its pre-images (which all have the same weight). This way we get a twin-free graph $\bar{H}$ such that $\hom(F, H) = \hom(F, \bar{H})$ for every graph F.

From now on, we will assume that H is twin-free.

The second reason giving non-realizable vectors in $\mathbb{R}^{q^k}$ takes more work to handle. For every $x \in \mathcal{G}_k$, the vector $(\hom_\varphi(x, H)\colon\ \varphi \in [q]^k)$ will be invariant under automorphisms of H (acting on index φ by right multiplication). It was proved in [5] that this is all:

Theorem 3.1. *If the model graph H is twin-free, then a vector $y \in \mathbb{R}^{[q]^k}$ is realizable as* $(\hom_\varphi(x, H)\colon\ \varphi \in [q]^k)$ *for some $x \in \mathcal{G}_k$ if and only if it is invariant under the automorphisms of H.*

We note that from this it is easy to determine the dimension of the algebras $\mathcal{G}_k/H$. Let $\mathrm{Aut}(H)$ denote the automorphism group of H.

Corollary 3.2. *If the model graph H is twin-free, then the dimension of $\mathcal{G}_k/H$ is equal to the number of orbits of* $\mathrm{Aut}(H)$ *on ordered k-tuples of nodes in H.*

3.5. Contractors and connectors

We can use Theorem 3.1 to construct some useful special elements in $\mathcal{G}_k$. It implies that there is an element $z \in \mathcal{G}_2$ such that

$$\hom_{ij}(z, H) = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases}$$

Such a quantum graph is called a *contractor*. The name comes from the following fact (which is easy to verify). For every 2-labeled graph F with no edge connecting the labeled nodes, let F' denote the 1-labeled graph that is obtained by identifying the labeled nodes. We extend this operation linearly over $\mathcal{G}_2$. Then for every 2-labeled quantum graph x,

$$\hom(xz, H) = \hom(x', H).$$

In [6] it was shown that for every weighted graph H on q nodes, there is a contractor that is a linear combination of series-parallel graphs with at most $(6q)^q$ nodes (we will only need the bound on the size).

Another useful construction will help us get rid of multiple edges. A k-labeled graph is *simple*, if it has no multiple edges, and its labeled nodes are independent. A k-labeled quantum graph is *simple*, if it is a combination of simple k-labeled graphs.

A *connector* is a 2-labeled quantum graph p that acts as a edge, i.e., $p \equiv K_2 \pmod{H}$. It was proved in [6] that for every weighted graph H, there exists a simple connector (note: K_2 is a connector, but it is not simple by our definition). In fact, this connector can be represented as a linear combination of paths with at most $q + 2$ nodes, labeled at their endpoints. Replacing each edge by a connector, we get:

Lemma 3.3. *Let x be any k-labeled quantum graph. Then there exists a simple k-labeled quantum graph y such that $x \equiv y \pmod{H}$.*

4. Proof of Theorem 2.2

Let $(G_1, G_2, \ldots)$ be a sequence of graphs such that $V(G_n] = [n]$ and

$$\frac{\hom(F, G_n)}{n^{|V(F)|}} \to \hom(F, H)$$

for every simple graph F (we shall see that we will use this condition only for a finite number of graphs F). Let G_n' denote the weighted graph obtained from G_n by weighting its nodes by $1/n$, so that now the condition can be written as

$$\hom\bigl(F, G_n'\bigr) \to \hom(F, H).$$

We will try to avoid confusion between G_n and H by denoting a typical node of H by i or j, and a typical node of G_n by u or v; a typical map into H will be denoted by φ, while a typical map into G_n (or G_n') will be denoted by η.

The graph H defines a seminorm $\|.\|$ on $\mathcal{G}_k$; the graph G_n' defines another seminorm, which we denote by $\|.\|_n$. Our condition implies that for every $x \in \mathcal{G}_k$,

$$\|x\|_n \to \|x\|.$$

4.1. More special quantum graphs

Recall that $\mathcal{G}_2$ has a contractor z for H. By Lemma 3.3, we may assume that z is simple. By replacing z by z^2 if necessary, we may assume that $\hom_\varphi(z, H) \geqslant 0$ for every graph H and $\varphi : [2] \to V(H)$.

By Theorem 3.1, there is a quantum graph $w \in \mathcal{G}_q$ such that

$$\hom_\phi(w, H) = \begin{cases} 1 & \text{if } \phi \text{ is bijective,} \\ 0 & \text{otherwise.} \end{cases}$$

By Lemma 3.3, we may assume that w is simple. Clearly $w^2 \equiv w \pmod H$, so we can replace w by w^2. Then $\hom_\varphi(w, H') \geqslant 0$ for every graph H' and every $\varphi : [q] \to V(H)$.

We define a number of special elements of $\mathcal{G}_q$. For $x \in \mathcal{G}_2$ and $y \in \mathcal{G}_k$, we say that y_1 is obtained from y by gluing x on nodes i and j ($i, j \in [k]$), if it is obtained by identifying the two labeled nodes of x with i and j, respectively; we keep the labeling as it was in y.

For every $i \in [q]$, we add a new isolated node to w, label it $q+1$, and glue a copy of z on $(i, q+1)$. Then we unlabel $q+1$, to get a quantum graph $w_i \in \mathcal{G}_q$.

For every $i, j \in [q]$, we add a new isolated node to w, label it $q+1$, and glue a copy of z on $(i, q+1)$ and another copy on $(j, q+1)$. Then we unlabel $q+1$, to get a quantum graph $w_{ij} \in \mathcal{G}_q$.

For every $i, j \in [q]$ and every bipartite graph F, we construct the disjoint union of w and F, and label the nodes of $U(F)$ by $q+1, \ldots, q+|U(F)|$ and the nodes of $W(F)$ by $q+|U(F)|+1, \ldots, q+|U(F)|+|W(F)|$. We glue a copy of z on each of the pairs $(i, q+1), \ldots, (i, q+|U(F)|)$ and also on each pair $(j, q+|U(F)|+1), \ldots, (j, q+|U(F)|+|W(F)|)$. Then we unlabel nodes $q+1, \ldots, q+|U(F)|+|W(F)|$, to get a quantum graph $w_{ij,F} \in \mathcal{G}_q$. We will only use this construction in two special cases: when $F = K_2$ and when $F = C_4$ (in both cases the bipartition is unique up to automorphisms).

We conclude this section with some properties of these quantum graphs under the map Ξ_q. We remarked before that $w \equiv w^2 \pmod H$. We also need that

$$\|w\|^2 = \sum_{\varphi : [q] \to [q]} \alpha_\varphi \hom_\varphi(w, H)^2 = \sum_{\varphi \in \mathrm{Aut}(H)} \alpha_\varphi = \bigl|\mathrm{Aut}(H)\bigr| \prod_{i \in [q]} \alpha_i. \tag{4}$$

We denote the number on the right-hand side by c. Similar arguments give the following equations:

$$\begin{aligned} \|w - w_1 - \cdots - w_q\| &= 0, \\ \|w_i - w_{ii}\| &= 0 \quad \bigl(\forall i \in [q]\bigr), \\ \|w_i - \alpha_i w\| &= 0 \quad \bigl(\forall i \in [q]\bigr), \\ \|w_{ij}\| &= 0 \quad \bigl(\forall i, j \in [q],\ i \neq j\bigr), \\ \bigl\|w_{ij,F} - \alpha_i^{|U(F)|} \alpha_j^{|W(F)|} \beta_{ij}^{|E(F)|} w\bigr\| &= 0 \quad \bigl(\forall i, j \in [q],\ \forall \text{ bipartite } F\bigr). \end{aligned}$$

(The last equation holds whether or not $i = j$.)

4.2. Constructing the partition

Now we look at the norm defined by G_n'. We know that

$$\|w\|_n \to \|w\| = c \quad (n \to \infty),$$

and similarly we get that as $n \to \infty$,

$$\begin{aligned} \|w - w_1 - \cdots - w_q\|_n &\to 0, \\ \|w_i - w_{ii}\|_n &\to 0 \quad \bigl(\forall i \in [q]\bigr), \\ \|w_i - \alpha_i w\|_n &\to 0 \quad \bigl(\forall i \in [q]\bigr), \\ \|w_{ij}\|_n &\to 0 \quad \bigl(\forall i, j \in [q],\ i \neq j\bigr), \\ \bigl\|w_{ij,F} - \alpha_i^{|U(F)|}\alpha_j^{|W(F)|}\beta_{ij}^{|E(F)|}w\bigr\|_n &\to 0 \quad \bigl(\forall i, j \in [q],\ \forall \text{ bipartite } F\bigr). \end{aligned}$$

So for a fixed $\varepsilon > 0$, we have

$$\bigl|\|w\|_n - c\bigr| < \varepsilon,$$

and so if $\varepsilon < c/2$, and n is large enough, we have $\|w\|_n > c/2$. On the other hand, we have

$$\|w\|_n^2 = \frac{1}{n^q} \sum_{\eta:[q]\to[n]} \hom_\eta\bigl(w, G'_n\bigr)^2,$$

and here every term is bounded by (3): $\hom_\eta(w, G'_n) \leqslant N(w)$. It follows that $N(w) \geqslant c/2$ and, for at least $c^2n^q/(8N(w)^2)$ maps η, we have $\hom_\eta(w, G'_n) \geqslant c/4$.

Now we look at the other special quantum graphs. We know that

$$\begin{aligned} &\|w - w_1 - \cdots - w_q\|_n^2 + \sum_{i=1}^q \|w_i - w_{ii}\|_n^2 + \sum_{i=1}^q \|w_i - \alpha_i w\|_n^2 + \sum_{1\leqslant i\neq j\leqslant q} \|w_{ij}\|_n^2 \\ &\quad + \sum_{1\leqslant i,j\leqslant q} \|w_{ij,K_2} - a_i\alpha_j\beta_{ij}w\|_n^2 + \sum_{1\leqslant i,j\leqslant q} \bigl\|w_{ij,C_4} - a_i^2\alpha_j^2\beta_{ij}^4 w\bigr\|_n^2 < \varepsilon \end{aligned}$$

if n is large enough. Let S denote this sum. We can write, for every quantum graph $x \in \mathcal{G}_q$,

$$\|x\|_n^2 = \frac{1}{n^q} \sum_{\eta:[q]\to[n]} \hom_\eta\bigl(x, G'_n\bigr)^2,$$

and so

$$\begin{aligned} S = \frac{1}{n^q} \sum_{\eta:[q]\to[n]} \Biggl(&\hom_\eta\bigl(w - w_1 - \cdots - w_q, G'_n\bigr)^2 \\ &+ \sum_i \hom_\eta\bigl(w_i - w_{ii}, G'_n\bigr)^2 + \sum_i \hom_\eta\bigl(w_i - \alpha_i w, G'_n\bigr)^2 \\ &+ \sum_{i\neq j} \hom_\eta\bigl(w_{ij}, G'_n\bigr)^2 + \sum_{i,j} \hom_\eta\bigl(w_{ij,K_2} - \alpha_i\alpha_j\beta_{ij}w, G'_n\bigr)^2 \\ &+ \sum_{i,j} \bigl(\hom_\eta\bigl(w_{ij,C_2} - \alpha_i^2\alpha_j^2\beta_{ij}^4 w, G'_n\bigr)\bigr)^2 \Biggr). \end{aligned}$$

Thus we can find an $\eta : [q] \to [n]$ such that

$$\hom_\eta\bigl(w, G'_n\bigr) \geqslant \frac{c}{4}, \tag{5}$$

and

$$\hom_\eta\big(w - w_1 - \cdots - w_q, G'_n\big)^2 < \varepsilon, \tag{6}$$

$$\sum_i \hom_\eta\big(w_i - w_{ii}, G'_n\big)^2 < \varepsilon, \tag{7}$$

$$\sum_i \hom_\eta\big(w_i - \alpha_i w, G'_n\big)^2 < \varepsilon, \tag{8}$$

$$\sum_{i\neq j} \hom_\eta\big(w_{ij}, G'_n\big)^2 < \varepsilon, \tag{9}$$

$$\sum_{i,j} \hom_\eta\big(w_{ij,K_2} - \alpha_i\alpha_j\beta_{ij} w, G'_n\big)^2 < \varepsilon, \tag{10}$$

$$\sum_{i,j} \hom_\eta\big(w_{ij,C_4} - \alpha_i^2\alpha_j^2\beta_{ij}^4 w, G'_n\big)^2 < \varepsilon. \tag{11}$$

We fix ε, n and this map η now. To simplify notation, we set $v_i = \eta(i)$, and for $u \in [n]$, we set $g_i(u) = \hom_{v_i u}(z, G'_n)$. Let $k_i(u) = 1$ if $g_i(u)$ is the largest among the numbers $g_j(u)$ ($j \in [q]$) and $k_i(u) = 0$ otherwise. (We break ties arbitrarily, so that $\sum_i k_i(u) = 1$ for all u.) We define a partition $[n] = V_1 \cup \cdots \cup V_q$ as follows: put u in V_i if $k_i(u) = 1$. We are going to prove that this partition satisfies the requirements of the theorem.

4.3. *A lemma about the partition*

The following lemma shows that, on the average, $g_i(u) \approx 1$ if $u \in V_i$ and $g_i(u) \approx 0$ otherwise.

Lemma 4.1.

$$\frac{1}{n}\sum_{u\in[n]}\sum_{i\in[q]}\big(g_i(u) - k_i(u)\big)^2 \leqslant \frac{256q\varepsilon}{c^2}.$$

Proof. We need an auxiliary function: For every $u \in [n]$ and $i \in [q]$, let

$$h_i(u) = \begin{cases} 1, & \text{if } g_i(u) \geqslant \frac{1}{2}, \\ 0, & \text{otherwise.} \end{cases}$$

We have

$$\begin{aligned} \hom_\eta\big(w_i - w_{ii}, G'_n\big)^2 &= \hom_\eta\big((w_i - w_{ii})^2, G'_n\big)^2 \\ &= \hom_\eta\big(w, G'_n\big)^2 \sum_{u\in[n]} \big(g_i(u) - g_i(u)^2\big)^2, \end{aligned}$$

and so it follows by (5) and (7) that

$$\frac{1}{n}\sum_{u\in[n]}\sum_{i\in[q]} g_i(u)^2\big(1 - g_i(u)\big)^2 \frac{16\varepsilon}{c^2}. \tag{12}$$

Similarly, (6) implies that

$$\frac{1}{n}\sum_{u\in[n]}\bigg(1 - \sum_{i\in[q]} g_i(u)\bigg)^2 \leqslant \frac{16\varepsilon}{c^2}. \tag{13}$$

Next we show that

$$\frac{1}{n}\sum_{u\in[n]}\sum_{i\in[q]}\big(g_i(u)-h_i(u)\big)^2\leqslant\frac{64\varepsilon}{c^2}. \tag{14}$$

Indeed, by the definition of $h_i(u)$, we have

$$\big(g_i(u)-h_i(u)\big)^2\leqslant 4g_i(u)^2\big(1-g_i(u)^2\big),$$

and so (14) follows by (12). We also claim that

$$\frac{1}{n}\sum_{u\in[n]}\sum_{i\in[q]}\big(h_i(u)-k_i(u)\big)^2\leqslant\frac{64q\varepsilon}{c^2}. \tag{15}$$

For a fixed $u\in[n]$, we have

$$\sum_{i\in[q]}\big(h_i(u)-k_i(u)\big)^2\leqslant\Big(1-\sum_{i\in[q]}h_i(u)\Big)^2,$$

since the sum on the left-hand side consists of $\sum_i h_i(u)$ terms of 1 if this sum is positive, and a single 1 if this sum is 0. So by (13) and (14),

$$\begin{aligned}
\frac{1}{n}\sum_{u\in[n]}\sum_{i\in[q]}\big(h_i(u)-k_i(u)\big)^2&\leqslant\frac{1}{n}\sum_{u\in[n]}\Big(1-\sum_{i\in[q]}h_i(u)\Big)^2\\
&\leqslant\frac{2}{n}\sum_{u\in[n]}\Big(1-\sum_{i\in[q]}g_i(u)\Big)^2+\frac{2}{n}\sum_{u\in[n]}\Big(\sum_{i\in[q]}\big(h_i(u)-g_i(u)\big)\Big)^2\\
&\leqslant\frac{2}{n}\sum_{u\in[n]}\Big(1-\sum_{i\in[q]}g_i(u)\Big)^2+\frac{2q}{n}\sum_{u\in[n]}\sum_{i\in[q]}\big(h_i(u)-g_i(u)\big)^2\\
&\leqslant\frac{32\varepsilon}{c^2}+\frac{32q\varepsilon}{c^2}\leqslant\frac{64q\varepsilon}{c^2}.
\end{aligned}$$

Now the lemma follows from (14) and (15):

$$\begin{aligned}
&\frac{1}{n}\sum_{u\in[n]}\sum_{i\in[q]}\big(g_i(u)-k_i(u)\big)^2\\
&\quad\leqslant\frac{2}{n}\sum_{u\in[n]}\sum_{i\in[q]}\big(g_i(u)-h_i(u)\big)^2+\frac{2}{n}\sum_{u\in[n]}\sum_{i\in[q]}\big(h_i(u)-k_i(u)\big)^2\\
&\quad\leqslant\frac{128\varepsilon}{c^2}+\frac{128q\varepsilon}{c^2}\leqslant\frac{256q\varepsilon}{c^2}.\qquad\square
\end{aligned}$$

4.4. The size of the classes

We prove that $|V_i|\approx\alpha_i n$. We first relate the size of V_i to w_i:

$$\begin{aligned}
\hom_\eta\big(w_i,G'_n\big)&=\frac{1}{n}\sum_{u\in[n]}\hom_\eta\big(w,G'_n\big)g_i(u)\\
&=\frac{c}{4n}\sum_{u\in[n]}k_i(u)+\frac{c}{4n}\sum_{u\in[n]}\big(g_i(u)-k_i(u)\big)^2
\end{aligned}$$

$$= \frac{c}{4n}|V_i| + R,$$

where the error term R satisfies

$$R^2 = \left(\frac{c}{4n}\sum_{u\in[n]}\big(g_i(u) - k_i(u)\big)\right)^2 \leqslant \frac{c^2}{16n}\sum_{u\in[n]}\big(g_i(u) - k_i(u)\big)^2 \leqslant 16q\varepsilon$$

by Lemma 4.1. So

$$\left|\frac{4\hom_\eta(w_i, G'_n)}{c} - \frac{|V_i|}{n}\right| \leqslant \frac{16\sqrt{q\varepsilon}}{c}.$$

On the other hand, (8) gives that

$$\left|\hom_\eta\big(w_i, G'_n\big) - \alpha_i \hom_\eta\big(w, G'_n\big)\right| \leqslant \sqrt{\varepsilon},$$

and so

$$\left|\frac{4\hom_\eta(w_i, G'_n)}{c} - \alpha_i\right| \leqslant \frac{4\sqrt{\varepsilon}}{c}.$$

So

$$\left|\frac{|V_i|}{n} - \alpha_i\right| \leqslant \frac{16\sqrt{q\varepsilon}}{c} + \frac{4\sqrt{\varepsilon}}{c} \leqslant c_1\sqrt{\varepsilon},$$

where c_1 is independent of n and ε. This proves assertion (a) of Theorem 2.2.

4.5. *Quasirandomness of the parts*

The proofs of (b) and (c) are similar, and we only describe the proof of (c). Let $1 \leqslant i < j \leqslant q$. We start with expressing the edge-density (in G'_n) between V_i and V_j. We have

$$\begin{aligned}
\hom_\eta\big(w_{ij,K_2}, G'_n\big) &= \frac{1}{n^2}\hom_\eta\big(w, G'_n\big)\sum_{uv\in E(G'_n)} g_i(u)g_j(v)\\
&= \frac{c}{4n^2}\sum_{uv\in E(G'_n)} k_i(u)k_j(v) + \frac{c}{4n^2}\sum_{uv\in E(G'_n)}\big(g_i(u)g_j(v) - k_i(u)k_j(v)\big)\\
&= \frac{c}{4n^2}\big|E_{G'_n}(V_i, V_j)\big| + R.
\end{aligned}$$

We estimate the error term as follows:

$$\begin{aligned}
R &= \frac{c}{4n^2}\sum_{uv\in E(G'_n)}\big(g_i(u)g_j(v) - k_i(u)k_j(v)\big)\\
&= \frac{c}{4n^2}\sum_{uv\in E(G'_n)}\big(g_i(u) - k_i(u)\big)k_j(v) + \frac{c}{4n^2}\sum_{uv\in E(G'_n)} g_i(u)\big(g_j(v) - k_j(v)\big).
\end{aligned}$$

To estimate the first term, we use that $k_j(v) \in \{0, 1\}$ and Lemma 4.1:

$$\begin{aligned}
\left(\frac{c}{4n^2}\sum_{uv\in E(G'_n)}\big(g_i(u) - k_i(u)\big)k_j(v)\right)^2 &\leqslant \frac{c}{4n^2}\sum_{uv\in E(G'_n)}\big(g_i(u) - k_i(u)\big)^2 k_j(v)^2\\
&\leqslant \frac{c^2}{16n}\sum_{u\in[n]}\big(g_i(u) - k_i(u)\big)^2 \leqslant 16q\varepsilon.
\end{aligned}$$

Estimating the second term is analogous, except that we have to use that $|g_i(u)| \leqslant N(z)$, and so we get $N(z)^2 16q\varepsilon$. Thus

$$R \leqslant 4\big(N(z)+1\big)\sqrt{q\varepsilon}.$$

Thus

$$\left|\frac{4\,\hom_\eta(w_{ij,K_2}, G_n')}{c} - \frac{E_{G_n'}(V_i, V_j)}{n^2}\right| \leqslant \frac{4R}{c} \leqslant \frac{16(N(z)+1)\sqrt{q\varepsilon}}{c}.$$

On the other hand, (10) gives that

$$\left|\frac{4\,\hom_\eta(w_{ij,K_2}, G_n')}{c} - \alpha_i\alpha_j\beta_{ij}\right| \leqslant \frac{4\sqrt{\varepsilon}}{c},$$

and so

$$\left|\frac{E_{G_n'}(V_i, V_j)}{n^2} - \alpha_i\alpha_j\beta_{ij}\right| \leqslant c_3\sqrt{\varepsilon},$$

where c_3 is independent of n and ε. We can write this as

$$\left|\frac{E_{G_n'}(V_i, V_j)}{|V_i|\cdot|V_j|} - \frac{\alpha_i n}{|V_i|}\frac{\alpha_j n}{|V_j|}\beta_{ij}\right| \leqslant c_3\frac{n^2}{|V_i|\cdot|V_i|}\sqrt{\varepsilon}.$$

Since we already know that $|V_i|/n \to \alpha_i$, as $\varepsilon \to 0$ and $n \to \infty$, this proves that the edge-density between V_i and V_j tends to β_{ij}.

An analogous argument, based on (11), shows that the density of C_4 in the bipartite graph formed by the edges between V_i and V_j tends to β_{ij}^4. By Theorem 2.1, this proves (c), and completes the proof of the theorem.

4.6. Finiteness

Theorem 2.3 follows by looking at some details of the proof. For a fixed H, we only used that $\hom(x, G_n') \to \hom(x, H)$ for a finite number of quantum graphs: w^2, $(w_i - w_{ii})^2$, w_{ij}^2, etc. Expanding the squares, it suffices to know $\hom(x, G_n') \to \hom(x, H)$ for $x \in W$, where

$$W = \big\{w^2, w_i^2, w_i w_{ii}, w_i w, w_{ij}^2, w_{ij,F}^2, w_{ij,F} w\colon\ i, j \in [q],\ F \in \{K_2, C_4\}\big\}.$$

These quantum graphs were composed of copies of z, w, and edges. We can express z and w as linear combinations of ordinary 2-labeled and q-labeled graphs:

$$z = \sum_{i=1}^{a} \lambda_i A_i,$$

and

$$w = \sum_{i=1}^{b} \mu_i B_i,$$

where $A_1, \ldots, A_a$ is a basis of $\mathcal{G}_2/H$ and $B_1, \ldots, B_b$ is a basis of $\mathcal{G}_q/H$. Then each $x \in W$ can be written as a linear combination of ordinary q-labeled graphs, obtained by replacing each z by one of the A_i and each w be one of the B_i. This gives a finite number of ordinary graphs $F_1, \ldots, F_r$, and if $\hom(F_i, G_n') \to \hom(F_i, H)$ for $i = 1, \ldots, r$, then the proof works and proves that G_n has the structure in Theorem 2.2, and hence it is quasirandom with model H.

The argument above gives an explicit bound on the number r. We have $a \leqslant q^2$ and $b \leqslant q^q$, by Corollary 3.2. The largest number of copies of w and z used in the same graph in W is 4 w's and 16 z's in w_{ij,C_4} (remember, we started with squaring z and w). So this gives at most $a^{16}b^4$ different graphs. There are fewer than $5q^2$ quantum graphs in W, which gives $r < 5q^{20q}$.

We also need to bound the graphs F_i we need. By the argument above, each F_i is glued together from at most 16 of the graphs A_i and 4 of the graphs B_i, so the proof of Theorem 2.3 will be complete if we prove the following bound on the size of ordinary graphs that generate $\mathcal{G}_k/H$:

Theorem 4.2. *The algebra $\mathcal{G}_k/H$ is generated by ordinary simple k-labeled graphs with at most $k+(10q)^q$ nodes.*

Proof. The idea is simple: let F be any k-labeled graph, and let $J \subseteq \binom{V(F)\setminus[k]}{2}$ be any set of pairs of elements in $V(F)\setminus[k]$. Let $\mathcal{H}_J$ denote the set of maps $\phi: V(F) \to [q]$ for which $\phi(x)=\phi(y)$ for every $\{x,y\} \in J$, and let $\psi:[k] \to [q]$. Define

$$\hom_{J,\psi}(F,H) = \sum_{\substack{\phi \in \mathcal{H}_J \\ \phi \text{ extends } \psi}} \alpha_\phi \beta_\phi.$$

Furthermore, let $\mathcal{I}$ be the set of injective maps $\phi: V(F) \to [q]$, and

$$\operatorname{inj}_\psi(F,H) = \sum_{\substack{\phi \in \mathcal{I} \\ \phi \text{ extends } \psi}} \alpha_\phi \beta_\phi.$$

Then by inclusion–exclusion,

$$\operatorname{inj}_\psi(F,H) = \sum_J (-1)^{|J|} \hom_{J,\psi}(F,H).$$

Suppose that $|V(F)| > q$, then the left-hand side is 0, so we get that

$$\hom_\psi(F,H) = \sum_{J \neq \emptyset} (-1)^{|J|-1} \hom_{J,\psi}(F,H).$$

Now "essentially" we have

$$\hom_{J,\psi}(F,H) = \hom_\psi(F/J,H),$$

where F/J is obtained from F by identifying all pairs of nodes in J. Considering the quantum graph

$$X = \sum_{J \neq \emptyset} (-1)^{|J|-1} F/J,$$

we have

$$\hom_\psi(F,H) = \hom_\psi(x,H)$$

for every ψ, which means that $F \equiv x \pmod{H}$. Since each graph in the definition of x has fewer nodes than F, we are done by induction (it seems).

The trouble is that identifying nodes in F may create loops, multiple edges, and, most significantly, F/J will have nodeweights: let k_i denote the number of nodes of F mapped onto $i \in V(F/J)\setminus[k]$, then for every $\phi : V(F/J) \to [q]$, we have

$$\alpha_\phi = \prod_{i\in V(F/J)} \alpha_{\phi(i)}^{k_i},$$

which depends on these nodeweights.

The way out is that temporarily we allow k-labeled ordinary graphs F that have positive integer nodeweights (k_i: $i \in V(F)\setminus[k]$) (it is convenient to leave the labeled nodes alone), positive integer edgeweights m_{ij} ($i, j \in [q]$, $i \neq j$) and each node $i \in V(F)\setminus[k]$ may carry a loop with a positive integer weight m_{ii}. Let us call such an F a *decorated graph*. For a decorated k-labeled graph F, and map $\phi : [k] \to [q]$, we can define

$$\hom_\phi(F, H) = \sum_{\substack{\psi : V(F)\to[q] \\ \psi \text{ extends } \varphi}} \prod_{i\in V(F)\setminus[k]} \alpha_{\psi(i)}^{k_i} \prod_{ij\in E(F)} \beta_{\psi(i)\psi(j)}^{m_{ij}}.$$

We can now form the linear space $\mathcal{G}_k^*$ of formal linear combinations of decorated graphs, define product, inner product, and congruence modulo H in it, and factor out the kernel as before. The inclusion–exclusion argument above gives that

Lemma 4.3. *The algebra $\mathcal{G}_k^*/H$ is generated by k-labeled decorated quantum graphs with at most q unlabeled nodes.*

Next we show that we can get rid of the large weights.

Lemma 4.4. *Let F be a decorated k-labeled graph. Then F is congruent modulo H to a linear combination of decorated k-labeled graphs that are isomorphic to F but all nodeweights are at most q and all edgeweights are at most q^2.*

Proof. Let $u \in V(F)\setminus[k]$ have nodeweight $k_u > q$. Let $F^{(r)}$ denote the decorated k-labeled graph obtained from F by reducing the weight of u by r. Consider the polynomial

$$\prod_{i=1}^{q}(x-\alpha_i) = \sum_{j=0}^{q} a_j x^{q-j}.$$

Then for every $\phi : V(F) \to [q]$, we have

$$\sum_{j=0}^{q} a_j \hom_\phi\bigl(F^{(j)}, H\bigr) = \sum_{j=0}^{q} a_j \alpha_{\phi(u)}^{-j} \hom_\phi(F, H) = 0,$$

and so we also have for every $\psi : [k] \to [q]$

$$\sum_{j=0}^{q} a_j \hom_\psi\bigl(F^{(j)}, H\bigr) = 0.$$

Thus

$$\hom_\psi(F) = -\sum_{j=1}^{q} \hom_\psi\bigl(F^{(j)}\bigr) = \hom_\psi(x, H),$$

where $x = -\sum_{j=1}^{q} F^{(j)}$ is a quantum graph in which all the terms have smaller total weight. By induction, the lemma follows. If any of the edgeweights is larger than q^2, we argue similarly. □

To conclude, it suffices to prove

Lemma 4.5. *Every decorated k-labeled graph is congruent modulo H to a linear combination of undecorated k-labeled graphs with at most $k + (10q)^q$ nodes.*

Proof. By Lemma 4.4, we may assume that the given quantum graph F has nodeweights at most q and edgeweights at most q^2. Replace each unlabeled node u in F by a set $S_u = \{u_1, \dots, u_{k_u}\}$ of k_u nodes, and attach a contractor to u_1 and u_j for $j = 2, \dots, k_u$. For every edge uv of F, insert m_{uv} edges between the nodes in S_u arbitrarily. (We may be forced to create multiple edges and loops.) We can replace a loop at $u_j \in S_u$ by attaching both labeled nodes of a simple connector to u_j. (This may create a double edge in this connector.) We now get rid of the multiple edges by replacing them with a simple connector.

The number of nodes in the contractors is at most (number of nodes in F) $\times$ (maximum nodeweight) $\times$ (maximum number of nodes in component of the contractor), which is at most $q^2(6q)^q$. The number of nodes in the connectors coming from loops is at most $q \times q \times 2 \times q = 2q^3$. The number of nodes in the connectors coming from other edges is at most $\binom{q+2}{2} \times q \times q < q^4$. This proves the lemma. □

This completes the proof of Theorem 4.2. □

Acknowledgments

We are indebted to Christian Borgs, Jennifer Chayes, Mike Freedman, Monique Laurent, Lex Schrijver, Miki Simonovits, Joel Spencer, Balázs Szegedy, Gábor Tardos and Kati Vesztergombi for many valuable discussions and suggestions on the topic of graph homomorphisms, and to the anonymous referee for suggesting many improvements.

References

[1] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, K. Vesztergombi, Counting graph homomorphisms, in: M. Klazar, J. Kratochvil, M. Loebl, J. Matoušek, R. Thomas, P. Valtr (Eds.), Topics in Discrete Mathematics, Springer, 2006, pp. 315–371.
[2] C. Borgs, J. Chayes, L. Lovász, V.T. Sós, K. Vesztergombi, Convergent graph sequences I: Subgraph frequencies, metric properties, and testing; Convergent graph sequences II: Multiway cuts and statistical physics, Manuscript.
[3] F.R. Chung, R.L. Graham, R.K. Wilson, Quasi-random graphs, Combinatorica 9 (1989) 345–362.
[4] M. Freedman, L. Lovász, A. Schrijver, Reflection positivity, rank connectivity, and homomorphisms of graphs, J. Amer. Math. Soc. 20 (2007) 37–51.
[5] L. Lovász, The rank of connection matrices and the dimension of graph algebras, European J. Combin. 27 (2006) 962–970.
[6] L. Lovász, B. Szegedy, Contractors and connectors of graph algebras, J. Combin. Theory Ser. B 96 (2006) 933–957.
[7] M. Simonovits, V.T. Sós, Szemerédi's partition and quasirandomness, Random Structures Algorithms 2 (1991) 1–10.
[8] M. Simonovits, V.T. Sós, Hereditary extended properties, quasi-random graphs and induced subgraphs, Combin. Probab. Comput. 12 (2003) 319–344.
[9] M. Simonovits, V.T. Sós, Hereditarily extended properties, quasi-random graphs and not necessarily induced subgraphs, Combinatorica 17 (1997) 577–596.
[10] A. Thomason, Pseudorandom graphs, in: Random Graphs '85, in: North-Holland Math. Stud., vol. 144, North-Holland, Amsterdam, 1987, pp. 307–331.