Bipartite graphs of approximate rank 1.

W. T. Gowers

§1. Introduction.

Quasirandomness is a central concept in graph theory, and has played an important part in arithmetic combinatorics as well. Roughly speaking, a notion of quasirandomness for a mathematical object such as a graph or a subset of a finite Abelian group is a property P such that any object that has P will automatically have many other properties that a randomly chosen object would be expected to have. For instance, a graph G with n vertices and $p\binom{n}{2}$ edges is called c-quasirandom if there are at most $(p^4+c)n^4$ ordered 4-cycles—that is, quadruples (x,y,z,w) of vertices such that xy, yz, zw and xw are all edges of G. If this property holds, then it can be shown that every set A of vertices in G spans approximately $p\binom{|A|}{2}$ edges, just as we would expect if G were a random graph with edge-probability p.

This fact and many others like it go back to papers of Thomason [T] and Chung, Graham and Wilson [CGW]. (See also the fundamental papers of Alon [A] and Alon and Milman [AM].) In this paper we shall consider a generalization of the notion of a quasirandom graph to what we call a bipartite graph *of approximate rank 1.* A bipartite graph has this property if its bipartite adjacency matrix is a quasirandom perturbation of a rank-1 matrix, or equivalently, if the graph behaves like a random bipartite graph with the given degree sequence. To get some idea of what these properties mean, let us take a bipartite graph G with vertex sets X and Y, of cardinalities m and n, respectively, and let us write $d_1(x)$ for the degree of a vertex $x\in X$ and $d_2(y)$ for the degree of a vertex in Y. If we choose large subsets $A\subset X$ and $B\subset Y$, then how many edges do we expect there to be between A and B? Well, each vertex $x\in A$ has a neighbourhood of size $d_1(x)$ in Y, and if we choose a random neighbour of x then we "expect" that its distribution in Y will be roughly proportional to the degree sequence in Y. In particular, the probability that a random neigbour of x belongs to B should be approximately $|E(G)|^{-1}\sum_{y\in B}d_2(y)$. (Note that $|E(G)|$, the number of edges of G, is equal to $\sum_{y\in Y}d_2(y)$.) Therefore, we "expect" the number of edges between A and B to be about $\sum_{x\in A}d_1(x)|E(G)|^{-1}\sum_{y\in B}d_2(y)$. Writing $\delta_1(x)$ and $\delta_2(y)$ for the *normalized degrees* $d_1(x)/|Y|$ and $d_2(y)/|X|$, and $\beta=|E(G)|/|X||Y|$ for the density of G, we can rewrite this as $\beta^{-1}\sum_{x\in A}\sum_{y\in B}\delta_1(x)\delta_2(y)$.

If the number of edges between A and B really is always about $\beta^{-1}\sum_{x\in A}\sum_{y\in B}\delta_1(x)\delta_2(y)$, then there is a certain sense, familiar from the conventional theory of quasirandom graphs, in which the adjacency matrix of G is approximated by the rank-1 matrix $\beta^{-1}\delta_1 \otimes \delta_2$. (This is notation for the function from $X \times Y$ to $\mathbb{R}$ that takes the value $\beta^{-1}\delta_1(x) \otimes \delta_2(y)$ at (x, y).) Writing $G(x, y)$ for the adjacency matrix, we shall prove that a certain graph-theoretic property is equivalent to the statement that the function $f = G - \beta^{-1}\delta_1 \otimes \delta_2$ is quasirandom in a sense that is explained in [G3] and which we shall recall in the next section. We can then use known characterizations of quasirandom functions to deduce related characterizations of graphs with approximate rank 1.

This is not the first time that quasirandom graphs with prescribed degree sequences have been considered: indeed, Chung and Graham have a paper on the subject [CG]. However, they look at graphs rather than bipartite graphs. One can convert any graph into a bipartite graph by doubling up its vertex set, so our results imply some characterizations of the class of graphs considered by Chung and Graham (we have not tried to deduce all their results, or even thought about whether this can be done). However, there is no obvious way of deducing our results from theirs, since a rank-1 bipartite graph has rank 2 if it is considered as a graph. (In case this remark is unclear, if G is a bipartite graph with vertex sets X and Y, then we use the term "adjacency matrix" for the obvious corresponding function defined on $X \times Y$. However, if we think of G as a graph with vertex set $X \cup Y$ that just happens to be bipartite, then its adjacency matrix is a function defined on $(X \cup Y)^2$. This adjacencey matrix will contain two blocks, each a copy of the bipartite adjacency matrix.)

The motivation for our results is not generalization for the sake of it, but the fact that graphs of approximate rank 1 arise naturally in the context of a proof in [G2] of the Balog-Szemerédi theorem [BS]. In the fourth section of the paper, we shall prove that if G is any dense bipartite graph with vertex sets X and Y, then X has a large subset X' such that the induced subgraph $G(X', Y)$ of G has approximate rank 1. This argument can be iterated (the next step would be to apply it to the graph $G(X \setminus X', Y)$), and this results in a "one-sided" regularity lemma for G: we can partition the vertex set X into not too many sets $X_1, \ldots, X_k$ in such a way that almost all of the graphs $G(X_i, Y)$ are "nice" in the sense that they have approximate rank 1.

Many of the results we state are more like templates: they illustrate techniques of proof that may have to be slightly modified for applications. In the fifth section, we give an example of this. We apply results from the earlier sections, some directly and some after appropriate modification, in order to obtain a new proof of a theorem of Green and

Konyagin in additive combinatorics. Our bound is weaker than theirs, but the proof has certain interesting features and gives an idea of how the methods of this paper can be used.

2. Basic facts about quasirandom functions.

In this section we briefly review the properties we shall need from quasirandom functions. As is customary, we present the definition as a theorem: the theorem states that certain properties of a function are equivalent, and the function is called quasirandom if it has one of these properties. As is also customary, we state these properties in terms of constants c_i, and when we say that a property involving c_i implies a property involving c_j we mean that for every $c_j > 0$ there exists $c_i > 0$ such that the implication is valid. Actually, the dependences between the constants will be very important to us later, so we shall be explicit about them after we have stated the theorem in a more qualitative way.

Theorem 1.1. *Let X and Y be sets of sizes m and n respectively and let $f : X \times Y \to [-1,1]$. Then the following statements are equivalent.*

(i) $\sum_{x,x'\in X}\sum_{y,y'\in Y} f(x,y)f(x',y)f(x,y')f(x',y') \leqslant c_1 m^2 n^2$.

(ii) For any pair of functions $u : X \to [-1,1]$ and $v : Y \to [-1,1]$ we have the inequality $\left|\sum_{x,y} f(x,y)u(x)v(y)\right| \leqslant c_2 mn$.

(iii) For any pair of sets $X' \subset X$ and $Y' \subset Y$ we have the inequality $\left|\sum_{x\in X'}\sum_{y\in Y'} f(x,y)\right| \leqslant c_3 mn$.

Moreover, if $\sum_{x,y} f(x,y) = 0$ then they are also equivalent to the following further statement.

(iv) For any pair of sets $X' \subset X$ and $Y' \subset Y$ we have the inequality $\sum_{x\in X'}\sum_{y\in Y'} f(x,y) \leqslant c_4 mn$.

We shall say that f is *c-quasirandom* if it has property (i) above with $c_1 = c$.

A proof of this result can be found in [G3 §3], where the following dependences are obtained between the constants c_i. If (i) holds, then (ii) holds with $c_2 = c_1^{1/4}$. It follows that (iii) and (iv) also hold with constant $c_1^{1/4}$. Moreover, these implications are valid for an arbitrary real-valued function f, and not just one that takes values in $[-1,1]$. In the opposite direction, if (iv) and the extra hypothesis on f hold, then (ii) holds with $c_2 = 12c_4$. If (iii) holds, then (ii) holds with $c_2 = 4c_3$. Finally, if (ii) holds, then (i) holds with $c_1 = c_2$.

The expression in (i) can be used to define a norm on functions $f : X \times Y \to \mathbb{R}$, which we shall denote by $\|f\|_{U_2}$ since it is closely related to the so-called U_2-norm for functions defined on finite Abelian groups. Let us briefly see why this is so. We shall prove the result for complex-valued functions.

Lemma 2.2. *Let X and Y be finite sets, and for any four functions $f_1, \dots, f_4$ from $X \times Y$ to $\mathbb{C}$ define $[f_1, f_2, f_3, f_4]$ to be*

$$\mathbb{E}_{x,x' \in X} \mathbb{E}_{y,y' \in Y} f_1(x,y) \overline{f_2(x,y') f_3(x',y)} f_4(x',y') \ .$$

Define also $\|f\|_{U_2}$ to be $[f,f,f,f]^{1/4}$. Then we have the inequality

$$[f_1, f_2, f_3, f_4] \leqslant \|f_1\|_{U_2} \|f_2\|_{U_2} \|f_3\|_{U_2} \|f_4\|_{U_2} \ .$$

Proof. This is a standard application of the Cauchy-Schwarz inequality. Observe that the left hand side can be written

$$\mathbb{E}_{x,x'} \Big(\mathbb{E}_y f_1(x,y) \overline{f_3(x',y)} \Big) \Big(\mathbb{E}_{y'} \overline{f_2(x,y')} f_4(x',y') \Big) \ .$$

Applying the Cauchy-Schwarz inequality, we deduce that

$$[f_1, f_2, f_3, f_4] \leqslant [f_1, f_1, f_3, f_3]^{1/2} [f_2, f_2, f_4, f_4]^{1/2} \ .$$

The same argument, but interchanging the roles of x and y, shows that

$$[f_1, f_2, f_3, f_4] \leqslant [f_1, f_2, f_1, f_2]^{1/2} [f_3, f_4, f_3, f_4]^{1/2} \ .$$

Applying the second inequality to the two terms on the right of the first then proves the lemma. □

Lemma 2.3. *Let X and Y be finite sets. Then the function $\|.\|_{U_2}$ defined in the previous lemma is a norm on the space $\mathbb{C}^{X \times Y}$.*

Proof. It is easy to check that $[f,f,f,f]$ is non-negative, and zero only when f is zero. Indeed,

$$[f,f,f,f] = \mathbb{E}_{x,x'} \Big| \mathbb{E}_y f(x,y) \overline{f(x',y)} \Big|^2 \ .$$

This is an expectation over pairs (x,x') of terms that are real and non-negative. Moreover, if $f(u,y) \neq 0$ for some u, y, then the term corresponding to the pair (u,u) is strictly positive. Obviously $\|\lambda f\| = |\lambda| \|f\|$ for any $\lambda \in \mathbb{C}$, so it remains to check the triangle inequality. But

$$\|f+g\|^4 = [f+g, f+g, f+g, f+g] \ .$$

Since $[f_1, f_2, f_3, f_4]$ is "quadrilinear" (the inverted commas are there since there are complex conjugates to think about, but the word "quadrisesquilinear" is not exactly pretty), the term on the right splits up as a sum of 16 such terms, with each f_i equal to f or g. Applying Lemma 2.2 to each of these terms and then using the binomial theorem, we deduce easily that

$$[f + g, f + g, f + g, f + g] \leqslant (\|f\| + \|g\|)^4 \ .$$

This proves the lemma. □

Thus, a quasirandom function is one with small U_2-norm. We have chosen the "correct" normalization for the U_2-norm, which is the one that has order of magnitude 1 for dense graphs regardless of the sizes of X and Y. Indeed, it is slightly more natural (though completely equivalent to the definition we gave above) to define f to be c-quasirandom if it satisfies the inequality

$$\mathbb{E}_{x,x' \in X} \mathbb{E}_{y,y' \in Y} f(x, y) f(x, y') f(x', y) f(x', y') \leqslant c \ .$$

A further characterization of quasirandom functions can be read out of a discussion in [G4]. If X and Y are sets of size m and n, respectively, and $f : X \times Y \to \mathbb{R}$ is any function, then f can be written in the form $\sum_{i=1}^{t} \lambda_i u_i \times v_i$, where $t = \min\{m, n\}$ and (u_i) and (v_i) are orthonormal sequences. This is known as the (or rather a) *singular-value decomposition* of f, and it is the non-symmetric analogue of the spectral decomposition of a symmetric matrix. We shall refer to the λ_i as the *singular values* of f. A proof that the singular-value decomposition exists is given as Theorem 2.6 in [G4] (but the result itself goes back to the 19th century). From the proof we have a fact that will be used later: that the largest modulus of any λ_i is equal to the largest inner product of f with any function $u \otimes v$ such that $\|u\|_2 = \|v\|_2 = 1$.

We remark that we are again using the uniform probability measures on X and Y to define our inner products. That is, when we say that $\|u_i\|_2 = 1$, we mean that $\big(\mathbb{E}_{x \in X} u_i(x)^2\big)^{1/2} = 1$. More generally, if u and u' are two functions from X to $\mathbb{C}$, we define $\langle u, u' \rangle$ to be $\mathbb{E}_{x \in X} u(x)\overline{u'(x)}$, with similar definitions for Y. Similar definitions also apply to functions defined on $X \times Y$. For example, if $f : X \times Y \to \mathbb{C}$, then we define $\|f\|_2$ to be $\big(\mathbb{E}_{x \in X} \mathbb{E}_{y \in Y} |f(x, y)|^2\big)^{1/2}$.

The following result, which is easy to prove, is very similar to Lemma 2.8 of [G4]. We give the proof in the complex case, but will apply it only in the real case.

Lemma 2.4. *Let X and Y be sets of size m and n, respectively, and let $f : X \times Y \to \mathbb{C}$ be a function with a singular-value decomposition $\sum_{i=1}^t \lambda_i u_i \otimes v_i$. Then $\sum_i |\lambda_i|^2 = \|f\|_2^2$ and $\sum_i |\lambda_i|^4 = \|f\|_{U_2}^4$.*

Proof. These two statements are both proved by setting $f = \sum_i \lambda_i u_i \otimes v_i$, expanding out the right-hand sides of the identities to be proved, and exploiting the orthnormality of the sequences (u_i) and (v_i). We shall abbreviate $\mathbb{E}_{x\in X}$ and $\mathbb{E}_{y\in Y}$ by $\mathbb{E}_x$ and $\mathbb{E}_y$, respectively. For the first identity,

$$\begin{aligned}\|f\|_2^2 &= \sum_{i=1}^t \sum_{j=1}^t \lambda_i \overline{\lambda_j} \mathbb{E}_x \mathbb{E}_y u_i(x) v_i(y) \overline{u_j(x) v_j(y)} \\ &= \sum_{i=1}^t \sum_{j=1}^t \lambda_i \overline{\lambda_j} \langle u_i, u_j \rangle \langle v_i, v_j \rangle \\ &= \sum_{i=1}^t |\lambda_i|^2 \ ,\end{aligned}$$

and for the second,

$$\begin{aligned}\|f\|_{U_2}^4 &= \sum_{i,j,k,l} \lambda_i \overline{\lambda_j \lambda_k} \lambda_l \mathbb{E}_{x,x'} \mathbb{E}_{y,y'} u_i(x) v_i(y) \overline{u_j(x) v_j(y') u_k(x') v_k(y)} u_l(x') v_l(y') \\ &= \sum_{i,j,k,l} \lambda_i \overline{\lambda_j \lambda_k} \lambda_l \langle u_i, u_j \rangle \langle v_i, v_k \rangle \langle u_l, u_k \rangle \langle v_l, v_j \rangle \\ &= \sum_{i=1}^t |\lambda_i|^4 \ ,\end{aligned}$$

since if a term is non-zero then we must have $i = j$, $i = k$, $k = l$ and $j = l$, and then all four inner products are 1. □

From this we obtain the following extra characterization of quasirandom functions.

Theorem 2.5. *Let X and Y be sets of size m and n, respectively, and let $f : X \times Y \to \mathbb{C}$ be a function such that $\|f\|_2 \leqslant 1$. Let $\lambda_1, \ldots, \lambda_k$ be the singular values of f. Then the following three statements are equivalent.*

(i) f is c_1-quasirandom.

(ii) $\sum_i |\lambda_i|^4 \leqslant c_1$.

(iii) $\max_i |\lambda_i| \leqslant c_2$.

Proof. As we have remarked, f is c_1-quasirandom if and only if $\|f\|_{U_2}^4 \leqslant c_1$, so the equivalence between (i) and (ii) follows straight from the second part of Lemma 2.4.

It is trivial that $\sum_i |\lambda_i|^4 \leqslant \max_i |\lambda_i|^2 \sum_i |\lambda_i|^2$. Therefore, if (iii) holds, then $\sum_i |\lambda_i|^4 \leqslant c_2^2$, by the first part of Lemma 2.4 and our assumption that $\|f\|_2 \leqslant 1$. Conversely, if (ii) holds then we clearly have $\max_i |\lambda_i|^4 \leqslant c_1$, so (iii) holds with $c_2 = c_1^{1/4}$. □

In practice, we shall apply Theorem 2.5 to functions f that map into the interval $[-1,1]$, so that it will be obvious that the L_2-estimate holds.

The next lemma gives us some information under weaker hypotheses than those of Lemma 2.4.

Lemma 2.6. *Let X and Y be two finite sets, let $f : X \times Y \to \mathbb{C}$ be a function and suppose that $f = \sum_{i=1}^t \lambda_i u_i \otimes v_i$, with (u_i) an orthonormal sequence and (v_i) an arbitrary sequence of vectors of L_2-norm 1. Then $\sum_i |\lambda_i|^2 = \|f\|_2^2$ and $\sum_i |\lambda_i|^4 \leqslant \|f\|_{U_2}^4$.*

Proof. We shall make use of the calculations in Lemma 2.4. For the first statement, notice that all we use about the sequences (u_i) and (v_i) is the fact that $\delta_{ij} = \langle u_i, u_j\rangle\langle v_i, v_j\rangle$, which is true under our weaker hypotheses. (There is no mystery about this: if $\langle u, u'\rangle = 0$ then it is easy to check that $\langle u \otimes v, u' \otimes v'\rangle = 0$ for any two vectors v and v'.)

For the second statement, notice that under the weaker hypotheses we can still use the orthonormality of the sequence (u_i) to replace the last line of the calculation by $\sum_{i=1}^t \sum_{k=1}^t |\lambda_i|^2|\lambda_k|^2\langle v_i, v_k\rangle\langle v_k, v_i\rangle$, which equals $\sum_{i=1}^t \sum_{k=1}^t |\lambda_i|^2|\lambda_k|^2|\langle v_i, v_k\rangle|^2$. Since all terms are non-negative, this double sum is at least as big as its diagonal, which is $\sum_{i=1}^t |\lambda_i|^4$ because $\langle v_i, v_i\rangle = 1$ for every i. □

We finish this section with a further lemma about U_2-norms. It tells us that if f and g are close in the U_2 norm, then for most pairs $x, x \in X$ the numbers $\mathbb{E}_y f(x,y)f(x',y)$ and $\mathbb{E}_y g(x,y)g(x',y)$, which we think of as "joint degrees" in the two "graphs" f and g, are approximately equal.

Lemma 2.7. *Let X and Y be finite sets and let f and g be functions from $X \times Y$ to $\mathbb{C}$. Then*

$$\mathbb{E}_{x,x'}\Big|\mathbb{E}_y\Big(f(x,y)\overline{f(x',y)} - g(x,y)\overline{g(x',y)}\Big)\Big|^2 \leqslant \|f-g\|_{U_2}^2\big(\|f\|_{U_2} + \|g\|_{U_2}\big)^2 .$$

Proof. Let $h = f - g$. Then

$$\mathbb{E}_y\Big(f(x,y)\overline{f(x',y)} - g(x,y)\overline{g(x',y)}\Big) = \mathbb{E}_y f(x,y)\overline{h(x',y)} + E_y h(x,y)\overline{g(x',y)} .$$

By Minkowski's inequality,

$$\Big(\mathbb{E}_{x,x'}\Big|\mathbb{E}_y f(x,y)\overline{h(x',y)}+\mathbb{E}_y h(x,y)\overline{g(x',y)}\Big|^2\Big)^{1/2}$$
$$\leqslant \Big(\mathbb{E}_{x,x'}\Big|\mathbb{E}_y f(x,y)\overline{h(x',y)}\Big|^2\Big)^{1/2}+\Big(\mathbb{E}_{x,x'}\Big|\mathbb{E}_y h(x,y)\overline{g(x',y)}\Big|^2\Big)^{1/2} .$$

Now $\mathbb{E}_{x,x'}\Big|\mathbb{E}_y f(x,y)\overline{h(x',y)}\Big|^2$ is another way of writing $[f,f,h,h]$, to use the notation of Lemma 2.2. Therefore, by that lemma, it is at most $\|f\|_{U_2}^2\|h\|_{U_2}^2$. Similarly, $\mathbb{E}_{x,x'}\Big|\mathbb{E}_y h(x,y)\overline{g(x',y)}\Big|^2$ is at most $\|g\|_{U_2}^2\|h\|_{U_2}^2$. Putting all this information together gives the result. □

§3. Characterizing functions of approximate rank 1.

We are now ready to prove some results about functions from $X\times Y$ to $\mathbb{R}$ that can be approximated by functions of rank 1. To obtain results about bipartite graphs, all one then needs to do is restrict attention to functions that take the value 0 or 1 everywhere.

We shall generalize concepts such as normalized degrees and density from graphs to functions in the obvious way. That is, if $f:X\times Y\to\mathbb{R}$ and $x\in X$ then we let $\delta_1(x)=\mathbb{E}_y f(x,y)$. Similarly, if $y\in Y$ then $\delta_2(y)$ is defined to be $\mathbb{E}_x f(x,y)$. Finally, the density d of G is $\mathbb{E}_{x,y}f(x,y)$. (Now that we are using normalized degrees, it is more convenient to use the letter d rather than β.) This density does not have to lie in the interval $[0,1]$, but for Theorem ??, which is the main result of the section, and the result that we shall apply later on, it will. We shall also write $\delta_1(x,x')$ for the joint degree $\mathbb{E}_y\delta(x,y)\delta(x',y)$.

Lemma 3.1. *Let X and Y be finite sets and let $h:X\times Y\to\mathbb{C}$. Then*

$$\mathbb{E}_x\Big|\mathbb{E}_y h(x,y)\Big|^2\leqslant\|h\|_{U_2}^2 .$$

Proof. This is a simple application of the Cauchy-Schwarz inequality:

$$\begin{aligned}\Big(\mathbb{E}_x\Big|\mathbb{E}_y h(x,y)\Big|^2\Big)^2&=\Big(\mathbb{E}_{y,y'}\mathbb{E}_x h(x,y)\overline{h(x,y')}\Big)^2\\&\leqslant\mathbb{E}_{y,y'}\Big|\mathbb{E}_x h(x,y)\overline{h(x,y')}\Big|^2\\&=\mathbb{E}_{x,x'}\mathbb{E}_{y,y'}h(x,y)\overline{h(x,y')h(x',y)}h(x',y')\\&=\|h\|_{U_2}^4 ,\end{aligned}$$

which implies the result. □

Lemma 3.2. *If X and Y are finite sets and $u : X \to \mathbb{R}$ and $v : Y \to \mathbb{R}$, then $\|u \otimes v\|_{U_2} = \|u \otimes v\|_2$.*

Proof. We can write $u \otimes v = \lambda u' \otimes v'$ with $\lambda = \|u\|_2 \|v\|_2$ and $\|u'\|_2 = \|v'\|_2 = 1$. Then Lemma 2.4 tells us that $\|u \otimes v\|_{U_2}^4 = \lambda^4$, which proves this lemma. □

The next lemma tells us that if f is close to a function of rank 1, then that function is approximately unique in L_2. This is true whether "close" means close in L_2 or close in the U_2-norm. We state it non-symmetrically, but obviously the same is true with the roles of the us and vs exchanged.

Lemma 3.3. *Let X and Y be finite sets, let u_1 and u_2 be functions from X to $\mathbb{C}$ and let v_1 and v_2 be functions from Y to $\mathbb{C}$. Suppose that either $\|u_1 \otimes v_1 - u_2 \otimes v_2\|_2$ or $\|u_1 \otimes v_1 - u_2 \otimes v_2\|_{U_2}$ is at most $\epsilon \|u_1\|_2 \|v_1\|_2$ (which equals both $\epsilon \|u_1 \otimes v_1\|_2$ and $\epsilon \|u_1 \otimes v_1\|_{U_2}$)). Then there exists a complex number λ such that $\|u_1 - \lambda u_2\|_2 \leqslant \epsilon \|u_1\|_2$ and $\|v_1 - \lambda^{-1} v_2\|_2 \leqslant \epsilon \|v_1\|_2 / (1 - \epsilon)$.*

Proof. Choose λ so that $u_1 - \lambda u_2$ is orthogonal to u_2. We know that

$$u_1 \otimes v_1 - u_2 \otimes v_2 = (u_1 - \lambda u_2) \otimes v_1 + \lambda u_2 \otimes (v_1 - \lambda^{-1} v_2) \ .$$

By Lemma 2.8, both the L_2-norm and the U_2-norm of this function are at least $\|u_1 - \lambda u_2\|_2 \|v_1\|_2$, so if one of them is at most $\epsilon \|u_1\|_2 \|v_1\|_2$ then $\|u_1 - \lambda u_2\|_2 \leqslant \epsilon \|u_1\|_2$. It follows that $\|\lambda u_2\|_2 \geqslant (1 - \epsilon) \lambda \|u_1\|_2$.

Lemma 2.8 also tells us that both the L_2-norm and the U_2-norm of the function above are at least $\|\lambda u_2\|_2 \|v_1 - \lambda^{-1} v_2\|_2$. Therefore, if either of them is at most $\epsilon \|u_1\|_2 \|v_1\|_2$, then $\|u_1 - \lambda u_2\|_2 \leqslant \epsilon \|u_1\|_2 \|v_1\|_2 / \|u_2\|_2$, which is at most $\epsilon \|v_1\| / (1 - \epsilon)$. □

Notice that there is an obvious converse to Lemma 3.3. If such a λ exists, then from the fact that

$$u_1 \otimes v_1 - u_2 \otimes v_2 = (u_1 - \lambda u_2) \otimes v_1 + \lambda u_2 \otimes (v_1 - \lambda^{-1} v_2)$$

we have immediately that both the L_2-norm and the U_2-norm of $u_1 \otimes v_1 - u_2 \otimes v_2$ are at most

$$\|u_1 - \lambda u_2\|_2 \|v_1\|_2 + \|\lambda u_2\|_2 \|v_1 - \lambda^{-1} v_2\|_2 \ ,$$

which is at most $\epsilon \|u_1\|_2 \|v_1\|_2 + \epsilon(1 + \epsilon) \|u_1\|_2 \|v_1\|_2$, which equals $(2\epsilon + \epsilon^2) \|u_1\|_2 \|v_1\|_2$.

Lemma 3.4. *Let X and Y be finite sets and let $f : X \times Y \to \mathbb{C}$ be a function with singular-value decomposition $\sum_{i=1}^t \lambda_i u_i \otimes v_i$. Then*

$$\max_i |\lambda_i| = \max\{\langle f, u \otimes v\rangle : \|u\|_2 = \|v\|_2 = 1\} \ .$$

Proof. Let u and v be unit vectors. Then

$$\langle f, u \otimes v\rangle = \sum_{i=1}^t \lambda_i \langle u, u_i\rangle\langle v, v_i\rangle \ .$$

By Bessel's inequality and the Cauchy-Schwarz inequality, $\left|\sum_{i=1}^t \langle u, u_i\rangle\langle v, v_i\rangle\right| \leqslant 1$. It follows that the left-hand side is at most the right-hand side. For the reverse inequality, let i be such that $|\lambda_i|$ is maximized and set $u = u_i$ and $v = v_i$. □

We shall now give some characterizations of functions from $X \times Y$ to $\mathbb{C}$ that can be approximated in L_2 by functions of rank 1. Later we shall use them to help us prove facts about functions that can be approximated in the U_2-norm.

Theorem 3.5. *Let $f : X \times Y \to \mathbb{C}$ and let $\sum_{i=1}^t \lambda_i u_i \otimes v_i$ be the singular-value decomposition of f with $|\lambda_1| \geqslant |\lambda_2| \geqslant \ldots$. Then the following statements are equivalent.*

(i) There exist $u : X \to \mathbb{C}$ and $v : Y \to \mathbb{C}$ with $\|u\|_2 = \|v\|_2 = 1$, and a constant $\lambda \in \mathbb{C}$, such that $\|f - \lambda u \otimes v\|_2^2 \leqslant c_1\|f\|_2^2$.

(ii) $\sum_{i\geqslant 2} |\lambda_i|^2 \leqslant c_2\|f\|_2^2$.

(iii) $\|f\|_{U_2}^4 \geqslant (1 - c_3)\|f\|_2^4$.

(iv) $\mathbb{E}_{x,x'}\mathbb{E}_{y,y'}\left|f(x, y)f(x', y') - f(x, y')f(x', y)\right|^2 \leqslant c_4\|f\|_2^4$.

Proof. If (ii) holds, then we can let $u = u_1$, $v = v_1$ and $\lambda = \lambda_1$. Then $\|f - \lambda u \otimes v\|_2 = \sum_{i\geqslant 2} |\lambda_i|^2$, by Lemma 2.4, so (i) holds with $c_1 = c_2$.

If (i) holds, then $\|f\|_2 \leqslant |\lambda| + c_1^{1/2}\|f\|_2$. It follows that $\sum_{i=1}^t |\lambda_i|^2 \leqslant |\lambda|^2 + 2|\lambda|c_1^{1/2}\|f\|_2 + c_1\|f\|_2^2$. We can also deduce from (i) and the Cauchy-Schwarz inequality that $|\langle f, u \otimes v\rangle| \geqslant |\lambda| - c_1^{1/2}\|f\|_2$. Therefore, by Lemma 3.4, $|\lambda_1| \geqslant |\lambda| - c_1^{1/2}\|f\|_2$, which implies that $|\lambda_1|^2 \geqslant |\lambda|^2 - 2c_1^{1/2}|\lambda|\|f\|_2 + c_1\|f\|_2^2$. It follows that $\sum_{i\geqslant 2} |\lambda_i|^2 \leqslant 4c_1^{1/2}|\lambda|\|f\|_2$. Since also $|\lambda| \leqslant \|f\|_2(1 + c_1^{1/2})$, this is at most $4(c_1^{1/2} + c_1)\|f\|_2^2$. Therefore, (ii) holds with $c_2 = 4(c_1^{1/2} + c_1)$.

The remaining equivalences are even more straightforward. Suppose that (ii) holds. Note first that $\|f\|_{U_2} \leqslant \|f\|_2$ by Lemma 2.4 and the fact that the ℓ_4-norm is dominated by the ℓ_2-norm. We also have that

$$\|f\|_2^4 = \Big(|\lambda_1|^2 + \sum_{i\geqslant 2}|\lambda_i|^2\Big)^2 \leqslant \|f\|_{U_2}^4 + (2c_2 + c_2^2)\|f\|_2^4 \ .$$

Therefore, (iii) is true with $c_3 = 2c_2 + c_2^2$.

Now suppose that (iii) holds. Then

$$\Big(\sum_{i=1}^t |\lambda_i|^2\Big)^2 \leqslant (1-c_3)^{-1}\sum_{i=1}^t |\lambda_i|^4 \leqslant (1-c_3)^{-1}|\lambda_1|^2\Big(\sum_{i=1}^t |\lambda_i|^2\Big) \ .$$

It follows that $\sum_{i=1}^t |\lambda_i|^2 \leqslant (1-c_3)^{-1}|\lambda_1|^2$, and therefore that (ii) is true with $c_2 = (1-c_3)^{-1} - 1$.

Finally, we show that (iii) and (iv) are equivalent by showing that the left-hand side of (iv) is another way of writing $2(\|f\|_2^4 - \|f\|_{U_2}^4)$. Indeed, expanding out the square gives us two "outer" terms, $|f(x,y)|^2|f(x',y')|^2$ and $|f(x,y')|^2|f(x',y)|^2$, both of which have expectation $\|f\|_2^4$, and two "inner" terms, $f(x,y)\overline{f(x,y')f(x',y)}f(x',y')$ and its complex conjugate, both of which have expectation $\|f\|_{U_2}^4$. Thus, (iv) is saying that $\|f\|_{U_2}^4 \geqslant (1-c_4/2)\|f\|_2^4$. □

Let f be as above. Then we shall write ff^* for the function from X^2 to $\mathbb{C}$ defined by the formula $ff^*(x,x') = \mathbb{E}_y f(x,y)\overline{f(x',y)}$. If we think of f as a matrix, then ff^* is the product of f with its adjoint—hence the notation.

Lemma 3.6. *Let X and Y be finite sets, and let $f : X \times Y \to \mathbb{C}$ be a function with singular-value decomposition $\sum_{i=1}^t \lambda_i u_i \otimes v_i$. Then $ff^* = \sum_{i=1}^t |\lambda_i|^2 u_i \otimes \overline{u_i}$.*

Proof. Since $f(x,y) = \sum_{i=1}^t \lambda_i u_i(x)v_i(y)$, we have

$$ff^*(x,x') = \mathbb{E}_y \sum_{i=1}^t\sum_{j=1}^t \lambda_i\overline{\lambda_j}u_i(x)v_i(y)\overline{u_j(x')v_j(y)} \ .$$

But $\mathbb{E}_y v_i(y)\overline{v_j(y)} = \delta_{ij}$, so this equals $\sum_{i=1}^t |\lambda_i|^2 u_i(x)\overline{u_i(x')}$, which proves the lemma. □

The next result characterizes functions $f : X \times Y \to \mathbb{C}$ that are close in the U_2-norm to a function of rank 1. As we shall see, this is true if and only if ff^* is close to a rank-1 function in the L_2-norm.

Theorem 3.7. *Let X and Y be finite sets, and let $f : X \times Y \to \mathbb{C}$ be a function with singular-value decomposition $\sum_{i=1}^t \lambda_i u_i \otimes v_i$, with $|\lambda_1| \geqslant \ldots \geqslant |\lambda_t|$. Then the following statements are equivalent.*

(i) There exist $u : X \to \mathbb{C}$ and $v : Y \to \mathbb{C}$ with $\|u\|_2 = \|v\|_2 = 1$, and a constant $\lambda \in \mathbb{C}$, such that $\|f - \lambda u \otimes v\|_{U_2}^4 \leqslant c_1 \|f\|_{U_2}^4$.

(ii) $\sum_{i \geqslant 2} |\lambda_i|^4 \leqslant c_2 \|f\|_{U_2}^4$.

(iii) There exists a function $u : X^2 \to \mathbb{C}$ such that $\|u\|_2 = 1$, and a constant $\mu \in \mathbb{R}_+$, such that $\|ff^* - \mu u \otimes \overline{u}\|_2^2 \leqslant c_3 \|f\|_{U_2}^4$.

(iv) $\mathbb{E}_{x_1,x_2,x_3,x_4} \Big| ff^*(x_1,x_3) ff^*(x_2,x_4) - ff^*(x_1,x_4) ff^*(x_2,x_3) \Big|^2 \leqslant c_4 \|f\|_{U_2}^8$.

Proof. The first equivalence is similar to the corresponding equivalence in Theorem 3.5. As there, it is trivial that if (ii) holds then (i) holds with the same constant.

If (i) holds, then $\|f\|_{U_2} \leqslant |\lambda| + c_1^{1/4}\|f\|_{U_2}$, so $\sum_{i=1}^t |\lambda_i|^4 \leqslant (|\lambda| + c_1^{1/4}\|f\|_{U_2})^4$. Writing $g = f - \lambda u \otimes v$, we have that $|\langle g, u \otimes v\rangle| \leqslant \|g\|_2 \leqslant \|g\|_{U_2} \leqslant c_1^{1/4}\|f\|_{U_2}$ by the Cauchy-Schwarz inequality, the fact that the U_2-norm is dominated by the L_2-norm, and (i). Therefore, $|\langle f, u \otimes v\rangle| \geqslant |\lambda| - c_1^{1/4}\|f\|_{U_2}$. Therefore, by Lemma 3.4, $|\lambda_1| \geqslant |\lambda| - c_1^{1/4}\|f\|_{U_2}$, so $|\lambda_1|^4 \geqslant (|\lambda| - c_1^{1/4}\|f\|_{U_2})^4$. It follows that

$$\begin{aligned}\sum_{i \geqslant 2} |\lambda_i|^4 &\leqslant (|\lambda| + c_1^{1/4}\|f\|_{U_2})^4 - (|\lambda| - c_1^{1/4}\|f\|_{U_2})^4 \\ &= 3c_1^{1/4}\|f\|_{U_2}|\lambda|^3 + 3c_1^{3/4}\|f\|_{U_2}|\lambda| \ .\end{aligned}$$

But (i) also implies that $|\lambda| \leqslant (1 + c_1^{1/4})\|f\|_{U_2}$, so this implies (ii) with $c_2 = 3c_1^{1/4}(1 + c_1^{1/4})^3 + 3c_1^{3/4}(1 + c_1^{1/4})$. In particular, if $c_1 \leqslant 1$, then it implies (ii) with $c_2 = 30c_1^{1/4}$ and for sufficiently small c_1 it implies it with $c_2 = 4c_1^{1/4}$.

The equivalence of (ii) and (iii) follows easily from Theorem 3.5 and Lemma 3.6. Indeed, suppose first that (ii) is true and set $\mu_i = |\lambda_i|^2$. By Lemma 3.6, the singular-value decomposition of ff^* is $\sum_{i=1}^t \mu_i u_i \otimes \overline{u_i}$, and by Lemma 2.4 we know that $\|f\|_{U_2}^4 = \sum_{i=1}^t |\lambda_i|^4 = \sum_{i=1}^t \mu_i^2 = \|ff^*\|_2^2$. Therefore, (ii) is telling us that $\sum_{i \geqslant 2} \mu_i^2 \leqslant c_2 \|ff^*\|_2^2$, which says that $\|ff^* - \mu_1 u_1 \otimes \overline{u_1}\|_2^2 \leqslant c_2 \|ff^*\|_2^2 = c_2 \|f\|_{U_2}^4$.

Conversely, if (iii) is true, then the implication of (ii) from (i) in Theorem 3.5 tells us that $\sum_{i \geqslant 2} \mu_i^2 \leqslant 4(c_3^{1/2} + c_3)\|ff^*\|_2^2$, which says that $\sum_{i \geqslant 2} |\lambda_i|^4 \leqslant 4(c_3^{1/2} + c_3)\|f\|_{U_2}^4$.

Similarly, the equivalence of (ii) and (iv) is just the equivalence between (ii) and (iv) of Theorem 3.5, but applied to the function ff^*. □

As remarked earlier, one can deduce results about graphs from results about functions. If f is the characteristic function of a bipartite graph G with vertex sets X and Y, then $ff^*(x, x') = |N_x \cap N_{x'}|/|Y|$, which we denoted earlier by $\delta_1(x, x')$. The next theorem combines several of the results stated so far. The equivalences involving (ii) are "weak" equivalences, in that the dependences between the c_i involve the U_2-norm of G. We make this remark precise in the proof.

Corollary 3.8. *Let X and Y be finite sets and let G be a bipartite graph with vertex sets X and Y of size m and n. Then the following statements are equivalent.*

(i) There exist functions $u : X \to \mathbb{R}$ and $v : Y \to \mathbb{R}$ with $\|u\|_2 = \|v\|_2 = 1$ and a real number λ such that $\|G - \lambda u \otimes v\|_{U_2}^4 \leqslant c_1 \|G\|_{U_2}^4$.

(ii) There exist functions $u : X \to \mathbb{R}$ and $v : Y \to \mathbb{R}$ with $\|u\|_2 = \|v\|_2 = 1$ and a real number λ such that for every subset $X' \subset X$ and every subset $Y' \subset Y$ the number of edges from X' to Y' differs from $\lambda\Big(\sum_{x\in X'} u(x)\Big)\Big(\sum_{y\in Y'} v(y)\Big)$ by at most $c_2 mn$.

(iii) $\mathbb{E}_{x_1,x_2,x_3,x_4}\Big|\delta(x_1,x_3)\delta(x_2,x_4) - \delta(x_1,x_4)\delta(x_2,x_3)\Big|^2 \leqslant c_3\|G\|_{U_2}^4$.

Moreover, the equivalence between (i) and (ii) holds for the same pair of functions u and v.

Proof. If we set $f = G - \lambda u \otimes v$, then (i) is equivalent to saying that f is c_1-quasirandom. That is, it is equivalent to saying that f satisfies property (i) of Theorem 1.1. Property (ii) of this theorem is easily seen to be equivalent to property (iii) of Theorem 1.1. Therefore, the equivalence of (i) and (ii) follows immediately from the equivalence of (i) and (iii) in Theorem 1.1. The dependences in that theorem imply that if (i) holds then (ii) holds with constant $c_1^{1/4}\|G\|_{U_2}$ and that if (ii) holds then (i) holds with constant $4c_2\|G\|_{U_2}^{-4}$.

The equivalence of (i) and (iii) is just the equivalence of (i) and (iv) in Theorem 3.7 applied to the function G. □

Before the last main result of the section we need two further technical lemmas. They have very similar proofs, but not quite similar enough for it to be worth the effort to deduce them from a single more abstract lemma.

Lemma 3.9. *Suppose that X and Y are finite sets and that f and g are functions from $X \times Y$ to $\mathbb{C}$. Then*

$$\mathbb{E}_{x,x'}\Big|\mathbb{E}_y\big(f(x,y)\overline{f(x',y)} - g(x,y)\overline{g(x',y)}\big)\Big|^2 \leqslant \|f-g\|_{U_2}^2(\|f\|_{U_2} + \|g\|_{U_2})^2$$

Proof. Let $h = f - g$. Then

$$f(x,y)\overline{f(x',y)} - g(x,y)\overline{g(x',y)} = f(x,y)\overline{h(x',y)} + h(x,y)\overline{g(x',y)}\ .$$

By Minkowski's inequality, the square root of the quantity we wish to bound is therefore at most

$$\Big(\mathbb{E}_{x,x'}\Big|\mathbb{E}_y f(x,y)\overline{h(x',y)}\Big|^2\Big)^{1/2} + \Big(\mathbb{E}_{x,x'}\Big|\mathbb{E}_y h(x,y)\overline{g(x',y)}\Big|^2\Big)^{1/2}\ .$$

Now $\mathbb{E}_{x,x'}\left|\mathbb{E}_y f(x,y)\overline{h(x',y)}\right|^2 = [f,f,h,h]$, so by Lemma 2.2 it is at most $\|f\|_{U_2}^2\|h\|_{U_2}^2$. A similar assertion is true for the second term, and the result follows easily. $\square$

Lemma 3.10. *Suppose that Z is a finite set and that f and g are functions from Z to $\mathbb{C}$. Then*

$$\mathbb{E}_{z,z'}\left|f(z)\overline{f(z')} - g(z)\overline{g(z')}\right|^2 \leqslant \|f-g\|_2^2(\|f\|_2 + \|g\|_2)^2$$

Proof. Once again we apply the standard trick for showing that products are close, observing that

$$f(z)\overline{f(z')} - g(z)\overline{g(z')} = f(z)(\overline{f(z')} - \overline{g(z')}) + (f(z) - g(z))\overline{g(z')}\ .$$

Again let $h = f - g$. Let us regard both sides of the above identity as functions from Z^2 to $\mathbb{C}$ and apply Minkowski's inequality. It tells us that

$$\left(\mathbb{E}_{z,z'}\left|f(z)\overline{f(z')} - g(z)\overline{g(z')}\right|^2\right)^{1/2} \leqslant \left(\mathbb{E}_{z,z'}\left|f(z)\overline{h(z')}\right|^2\right)^{1/2} + \left(\mathbb{E}_{z,z'}\left|h(z)\overline{g(z')}\right|^2\right)^{1/2}\ .$$

The right-hand side is equal to $\|h\|_2(\|f\|_2 + \|g\|_2)$, so the result is proved. $\square$

The message of the next result is that if f is a function from $X \times Y$ to $\mathbb{R}_+$ and f can be approximated by a rank-1 function $\lambda u \otimes v$, then we can identify in a very simple way what u, v and λ are. (By Lemma 3.3, they are approximately unique up to multiplication of u by a scalar and v by the inverse of that scalar.) Indeed, u and v are proportional to the (normalized) degrees of the vertices in X and Y, respectively. To obtain the constant of proportionality, one looks at $d = \mathbb{E}_{x,y}f(x,y)$. Of course, this will not work if $d = 0$, so the result is valid only when d is non-zero, and gets progressively weaker when d gets smaller.

It turns out that this problem persists to some extent even if f takes values that are real and positive (the case that will eventually interest us). When we prove that (i) implies (ii) below, we shall obtain a "weak" result, in that the constant c_2 that we obtain will depend not just on c_1 but also on the ratio $d^{-1}\|f\|_{U_2}$. After the proof, we shall give a simple example that demonstrates that this weakness is necessary.

Theorem 3.11. *Let X and Y be sets of size m and n, respectively, and let $f : X \times Y \to \mathbb{C}$. Then the following four statements are equivalent.*

(i) There exist functions $u : X \to \mathbb{C}$ and $v : Y \to \mathbb{C}$ such that $\|f - u \otimes v\|_{U_2}^4 \leqslant c_1\|f\|_{U_2}^4$.

(ii) $\|f - d^{-1}\delta_1 \otimes \delta_2\|_{U_2}^4 \leqslant c_2\|f\|_{U_2}^4$.

(iii) $\mathbb{E}_{x,x'}\Big|\delta_1(x,x') - (d^{-2}\|\delta_2\|_2^2)\delta_1(x)\delta_1(x')\Big|^2 \leqslant c_3\|f\|_{U_2}^4$.

(iv) $\mathbb{E}_{y,y'}\Big|\delta_2(y,y') - (d^{-2}\|\delta_1\|_2^2)\delta_2(y)\delta_2(y')\Big|^2 \leqslant c_4\|f\|_{U_2}^4$.

Proof. Trivially (ii) implies (i) with $c_1 = c_2$. To see that (i) implies (ii), we apply Lemma 3.1 to the function $h(x,y) = f(x,y) - u(x)v(y)$. Setting $\alpha = \mathbb{E}_x u(x)$ and $\beta = \mathbb{E}_y v(y)$, this tells us that

$$\mathbb{E}_x|\delta_1(x) - \beta u(x)|^2 \leqslant c_1^{1/2}\|f\|_{U_2}^2 \ .$$

That is, $\|\delta_1 - \beta u\|_2 \leqslant c_1^{1/4}\|f\|_{U_2}$. Similarly, $\|\delta_2 - \alpha v\|_2 \leqslant c_1^{1/4}\|f\|_{U_2}$. It follows that

$$\begin{aligned}\|\delta_1 \otimes \delta_2 - \alpha\beta u \otimes v\|_{U_2} &\leqslant \|\delta_1 \otimes (\delta_2 - \alpha v)\|_{U_2} + \|(\delta_1 - \beta u) \otimes \alpha v\|_{U_2} \\ &= \|\delta_1 \otimes (\delta_2 - \alpha v)\|_2 + \|(\delta_1 - \beta u) \otimes \alpha v\|_2 \\ &\leqslant c_1^{1/4}\|f\|_{U_2}\big(\|\delta_1\|_2 + \alpha\|v\|_2\big) \ .\end{aligned}$$

Now Lemma 3.1 applied to f tells us that $\|\delta_1\|_2 \leqslant \|f\|_{U_2}$, and $\alpha\|v\|_2 \leqslant \|u\|_2\|v\|_2 = \|u \otimes v\|_{U_2} \leqslant (1 + c_1^{1/4})\|f\|_{U_2}$. Therefore, we have shown that

$$\|\delta_1 \otimes \delta_2 - \alpha\beta u \otimes v\|_{U_2} \leqslant c^{1/4}(2 + c_1^{1/4})\|f\|_{U_2}^2 \ .$$

From Lemma 3.4 with u and v equal to the constant functions 1_X and 1_Y, respectively, and $f = h$, we find that $|\mathbb{E}_{x,y}h(x,y)| \leqslant \|h\|_{U_2}$. (This uses Lemma 2.4 as well.) We are assuming that $\|h\|_{U_2} \leqslant c_1^{1/4}\|f\|_{U_2}$. It follows that $|\alpha\beta - d| \leqslant c_1^{1/4}\|f\|_{U_2}$, and therefore that $\|\alpha\beta u \otimes v - du \otimes v\|_{U_2} \leqslant c_1^{1/4}\|f\|_{U_2}\|u\|_2\|v\|_2$, which is at most $c_1^{1/4}(1 + c_1^{1/4})\|f\|_{U_2}^2$.

Thus, by the triangle inequality,

$$\|\delta_1 \otimes \delta_2 - du \otimes v\|_{U_2} \leqslant c_1^{1/4}(3 + 2c_1^{1/4})\|f\|_{U_2}^2 \ ,$$

and by (i) and the triangle inequality again this tells us that

$$\|f - d^{-1}\delta_1 \otimes \delta_2\|_{U_2} \leqslant c_1^{1/4}\|f\|_{U_2} + d^{-1}c^{1/4}(3 + 2c_1^{1/4})\|f\|_{U_2}^2 \ .$$

Notice that the c_2 that we have obtained is equal to $(c_1^{1/4} + d^{-1}\|f\|_{U_2}c^{1/4}(3+2c_1^{1/4}))^4$, so it depends on the quantity $d^{-1}\|f\|_{U_2}$, as we warned that it would.

Now let us prove that (ii) implies (iii). We do this by applying Lemma 3.9 to the functions f and $d^{-1}\delta_1 \otimes \delta_2$, from which we learn that

$$\mathbb{E}_{x,x'}\Big|\delta_1(x,x') - (d^{-2}\|\delta_2\|_2^2)\delta_1(x)\delta_1(x')\Big|^2 \leqslant \|f - d^{-1}\delta_1 \otimes \delta_2\|_{U_2}^2(\|f\|_{U_2} + \|d^{-1}\delta_1 \otimes \delta_2\|_{U_2})^2 \ .$$

By (ii) and the triangle inequality, we know that $\|d^{-1}\delta_1 \otimes \delta_2\|_{U_2} \leqslant (1 + c_2^{1/4})\|f\|_{U_2}$. Therefore, the right-hand side of the above inequality is at most $c_2^{1/2}(2 + c_2^{1/4})^2\|f\|_{U_2}^4$, which establishes (iii) with $c_3 = c_2^{1/2}(2 + c_2^{1/4})^2$.

Finally, we prove that (iii) implies (i). Note first that $\delta_(x, x')$ is just a different notation for $ff^*(x, x')$. Therefore, property (iii) implies property (iii) of Theorem 3.7. We therefore have property (i) of that theorem (by the theorem) with constant $4(c_3^{1/2} + c_3)$, and hence property (i) of this theorem with the same constant.(A tiny extra ingredient is the observation that if f is real then its singular-value decomposition is real, so that the u, v and λ that come out of Theorem 3.7 are all real.)

Finally, notice that statements (i) and (ii) are symmetric in X and Y. Therefore, the equivalence with (iii) implies the equivalence with (iv) with exactly the same dependences. (Thus, interestingly, if you know that $\delta_1(x, x')$ tends to be roughly proportional to $\delta_1(x)\delta_1(x')$ then you know the same for δ_2. However, this is a "weak" implication.) □

To see why the weakness was necessary in the above result, consider the following example (or rather, class of examples). The basic idea behind it is to construct a function of the form $u \otimes v + u' \otimes v'$ in such a way that u, v, u' and v' are all positive, $\|u \otimes v\|_{U_2}$ is much bigger than $\|u' \otimes v'\|_{U_2}$, but $\mathbb{E}_{x,y}u(x)v(y)$ is much smaller than $\mathbb{E}_{x,y}u'(x)v'(y)$. This means that (i) is true, but (ii) is false.

To be more concrete, let $X_1 \subset X$ and $Y_1 \subset Y$ be subsets of density β, and let u and v be their characteristic functions. Let u' and v' be the characteristic functions of X and Y. Then let $f = u \otimes v + \beta^2 u' \otimes v'$. Now $\|u \otimes v\|_{U_2} = \|u\|_2\|v\|_2 = \beta$, and $\|\beta^2 u' \otimes v'\|_{U_2} = \beta^2$. Therefore, $\|f - u \otimes v\|_{U_2} \leqslant \beta\|u \otimes v\|_{U_2} \leqslant \beta(1 - \beta)^{-1}\|f\|_{U_2}$, so we have (i) with constant $\beta(1 - \beta)^{-1}$.

Now the density d of f is $2\beta^2$. If $x \in X_1$ and $y \in Y_1$ then $d^{-1}\delta_1(x)\delta_2(y) = d^{-1}(\beta + \beta^2)^2 \leqslant 1/2 + 3\beta$, whereas $f(x, y) = 1$. If $x \in X_1$ and $y \in Y_2$ then $d^{-1}\delta_1(x)\delta_2(y) = (\beta + \beta^2)\beta^2 \leqslant 2\beta^3$, and the same is true if $x \in X_2$ and $y \in Y_1$. In both these cases $f(x, y) = 0$. Finally, if $x \in X_2$ and $y \in Y_2$ then $d^{-1}\delta_1(x)\delta_2(y) = d^{-1}\beta^4 = \beta^2/2$.

We can split $g = f - d^{-1}\delta_1 \otimes \delta_2$ into four pieces of rank 1, by looking at its restrictions to the four sets $X_i \times Y_j$. The calculations above show that the part in $X_1 \times Y_1$ has U_2-norm at least $\beta(1/2 - 3\beta)$, since g is constant, and the constant is at least $1/2 - 3\beta$. Similarly, the parts in $X_1 \times Y_2$ and $X_2 \times Y_1$ have U_2-norm at most $2\beta^3$. Finally, the part in $X_2 \times Y_2$ has U_2-norm at most $\beta^2/2$, since g is constant and the constant lies between 0 and $\beta^2/2$. It follows that the U_2-norm of g is at least $\beta(1/2 - 3\beta) - 2\beta^3 - \beta^2/2$, which is at least $\beta/3$ if β is sufficiently small.

We know that $\|f\|_{U_2} \geqslant \beta$ (since u' and v' are positive), so this tells us that f is not close to $d^{-1}\delta_1 \otimes \delta_2$ in the U_2-norm. More precisely, (ii) is not true with a constant that tends to zero with β, while (i) is true with a constant that is roughly proportional to β. This shows that in order to obtain (ii), we need c_1 to be small in a way that depends not just on c_2 but also on the function f. In other words, the weak implication cannot be replaced by a strong one.

A special case that will interest us later is when a graph (or function) can be approximated by a function of the form $1 \otimes v$. The next theorem characterizes such functions in a way that is quite similar to the characterizations of Theorem 3.11, though we add one further characterization. However, for this theorem, perhaps surprisingly, all the implications are strong ones, except that the extra characterization (v) implies the others only weakly.

Theorem 3.12. *Let X and Y be finite sets and let $f : X \times Y \to \mathbb{R}_+$. Then the following statements are equivalent.*

(i) There exists a function $v : Y \to \mathbb{R}_+$ such that $\|f - 1 \otimes v\|_{U_2}^4 \leqslant c_1\|f\|_{U_2}^4$.

(ii) $\|f - 1 \otimes \delta_2\|_{U_2}^4 \leqslant c_2\|f\|_{U_2}^4$.

(iii) $\mathbb{E}_{x,x'}\Big|\delta_1(x, x') - \|\delta_2\|_2^2\Big|^2 \leqslant c_3\|f\|_{U_2}^4$.

(iv) $\mathbb{E}_{y,y'}\Big|\delta_2(y, y') - \delta_2(y)\delta_2(y')\Big|^2 \leqslant c_4\|f\|_{U_2}^4$.

(v) $\mathbb{E}_y\big|\mathbb{E}_x f(x, y)u(x)\big|^2 \leqslant c_5\|f\|_{U_2}^2\|u\|_2^2$ for every function $u : X \to \mathbb{R}$ of mean 0.

Proof. First let us show that (ii) and (iv) are exactly equivalent, by showing that their left-hand sides are equal. If we set $g(x, y) = f(x, y) - \delta_2(y)$, then the left-hand side of (ii) is equal to $\|g\|_{U_2}^4$, which is equal to $\mathbb{E}_{y,y'}|\mathbb{E}_x g(x, y)g(x, y')|^2$. But

$$\begin{aligned}\mathbb{E}_x g(x, y)g(x, y') &= \mathbb{E}_x f(x, y)f(x, y') - \mathbb{E}_x f(x, y)\delta_2(y') - \mathbb{E}_x f(x, y')\delta_2(y) + \delta_2(y)\delta_2(y') \\ &= \delta_2(y, y') - \delta_2(y)\delta_2(y')\ ,\end{aligned}$$

from which the equality follows.

The proof that (ii) implies (iii) is almost identical to the proof that (ii) implies (iii) in Theorem 3.11. Indeed, let us prove a more general result that implies them both. Suppose that $\|f - u \otimes v\|_{U_2}^4 \leqslant c_2\|f\|_{U_2}^4$. We can then apply Lemma 3.9 to the functions f and $u \otimes v$, concluding that

$$\mathbb{E}_{x,x'}\Big|\mathbb{E}_y\big(f(x, y)\overline{f(x', y)} - u(x)\overline{u(x')}|v(y)|^2\big)\Big|^2 \leqslant \|f - u \otimes v\|_{U_2}^2(\|f\|_{U_2} + \|u \otimes v\|_{U_2})^2\ .$$

By (ii) and the triangle inequality, $\|u\otimes v\|_{U_2}\leqslant(1+c_2^{1/4})\|f\|_{U_2}$, so

$$\mathbb{E}_{x,x'}\Big|\delta_1(x,x')-u(x)\overline{u(x')}\|v\|_2^2\Big|^2\leqslant c_2^{1/2}(2+c_2^{1/4})^2\|f\|_{U_2}^4\ .$$

Specializing to the case $u=1$ and $v=\delta_2$ tells us that (ii) implies (iii), and specializing to the case $u=\delta_1$ and $v=d^{-1}\delta_2$ tells us that (ii) implies (iii) in Theorem 3.11.

We finish the proof that all of (i), (ii), (iii) and (iv) are equivalent by proving that (iii) implies (i) and (i) implies (ii). The proofs are very similar to the corresponding proofs in Theorem 3.11 but slightly more complicated.

First, notice that (iii) implies property (iii) of Theorem 3.7 with the same constant c_3. It therefore gives us property (i) of Theorem 3.7 with constant $4(c_3^{1/2}+c_3)$, which we shall call c_6. Now the proof of that implication actually tells us slightly more—that $\|f-\lambda u\otimes v\|_{U_2}^4\leqslant c_6\|f\|_{U_2}^4$ where λ, u and v are the beginning of the singular-value decomposition of f. Since f is a non-negative function, λ, u and v can all be taken to be real and non-negative. It follows that $\|ff^*-\lambda^2u\otimes u\|_2^2\leqslant c_6\|f\|_{U_2}^4$ as well, since if $f=\sum_{i=1}^t\lambda_iu_i\otimes v_i$ with (u_i) and (v_i) both orthonormal sequences then $ff^*=\sum_{i=1}^t|\lambda_i|^2u_i\otimes\overline{u_i}$. But (iii) tells us that $\big\|ff^*-\|\delta_2\|_2^21\otimes 1\big\|_2^2\leqslant c_3\|f\|_{U_2}^4$, so it follows from Minkowski's inequality that $\big\|\lambda^2u\otimes u-\|\delta_2\|_2^21\otimes 1\big\|_2\leqslant(c_3^{1/2}+c_6^{1/2})\|f\|_{U_2}^2$.

We also know that $\lambda^2\|u\otimes u\|_2\geqslant\|ff^*\|_2-c_6^{1/2}\|f\|_{U_2}^2=(1-c_6^{1/2})\|f\|_{U_2}^2$. Therefore, $\big\|\lambda^2u\otimes u-\|\delta_2\|_2^21\otimes 1\big\|_2\leqslant(c_3^{1/2}+c_6^{1/2})(1-c_6^{1/2})^{-1}\lambda^2\|u\otimes u\|_2$. Let $c_7=(c_3^{1/2}+c_6^{1/2})(1-c_6^{1/2})^{-1}$. It follows from Lemma 3.3 that there is some constant function ρ such that $\|\lambda u-\rho\|_2\leqslant c_7\lambda\|u\|_2$. Therefore, $\|(\lambda u-\rho)\otimes v\|_{U_2}\leqslant c_7\lambda\|u\|_2\|v\|_2$. Hence, by the triangle inequality,

$$\|f-\rho\otimes v\|_{U_2}\leqslant c_6^{1/4}\|f\|_{U_2}+c_7\lambda\|u\|_2\|v\|_2\ .$$

We also know that $\|\lambda u\otimes v\|_{U_2}\leqslant(1+c_6^{1/4})\|f\|_{U_2}$, so this is at most $(c_6^{1/4}+c_7(1+c_6^{1/4}))\|f\|_{U_2}$. We have just established (i) with constant $c_1=\big(c_6^{1/4}+c_7(1+c_6^{1/4})\big)^4$.

If we now apply Lemma 3.1 to the function $f(x,y)-\rho v(y)$ (but with x and y interchanged) we find that $\mathbb{E}_y|\delta_2(y)-\rho v(y)|^2\leqslant\|f-\rho\otimes v\|_{U_2}^2$, which we know is at most $c_1^{1/2}\|f\|_{U_2}^2$. That is, $\|\delta_2-\rho v\|_2\leqslant c_1^{1/4}\|f\|_{U_2}$. (Here we are sometimes using ρ to stand for a constant function and sometimes to stand for the relevant constant.)

Now $\rho\otimes v=1\otimes\rho v$, so the above estimates give us

$$\begin{aligned}\|f-1\otimes\delta_2\|_{U_2}&\leqslant\|f-1\otimes\rho v\|_{U_2}+\|1\otimes(\rho v-\delta_2)\|_{U_2}\\&\leqslant c_1^{1/4}\|f\|_{U_2}+c_1^{1/4}\|f\|_{U_2}=2c_1^{1/4}\|f\|_{U_2}\ .\end{aligned}$$

This proves (ii) with constant $c_2=2c_1^{1/4}$.

To complete the proof of the lemma, we shall show that (iii) and (v) are equivalent. Since it is no harder, we shall do this for complex functions. Notice that property (v) is saying that if we associate with f a linear map from $L_2(X)$ to $L_2(Y)$ in the obvious way, then the restriction of this map to the set of functions with mean zero has norm at most $c_5^{1/2}\|f\|_{U_2}$. The proof that it follows from (iii) is the following simple calculation:

$$\begin{aligned}\mathbb{E}_y|\mathbb{E}_x u(x)f(x,y)|^2 &= \mathbb{E}_y\mathbb{E}_{x,x'}u(x)\overline{u(x')}f(x,y)\overline{f(x',y)}\\ &= \mathbb{E}_{x,x'}u(x)\overline{u(x')}\delta_1(x,x')\\ &= \mathbb{E}_{x,x'}u(x)\overline{u(x')}(\delta_1(x,x')-\alpha)\\ &\leqslant \Big(\mathbb{E}_{x,x'}|u(x)|^2|u(x')|^2\Big)^{1/2}\Big(\mathbb{E}_{x,x'}|\delta_1(x,x')-\alpha|^2\Big)^{1/2} .\end{aligned}$$

Here $\alpha=\|\delta_2\|_2^2$, the mean of $\delta_1(x,x')$. The first term in the final product equals $\|u\|_2^2$ and the second is at most $c_3^{1/2}\|f\|_{U_2}^2$ by hypothesis.

Finally, let us assume (v). We begin by deducing from it a statement about arbitrary functions u. Let $u:X\to\mathbb{C}$ be a function with mean τ and write $u(x)=\tau+v(x)$. Then

$$\mathbb{E}_{x,x'}\delta_1(x,x')u(x)\overline{u(x')}=\mathbb{E}_{x,x'}\delta_1(x,x')(\tau+v(x))(\overline{\tau}+\overline{v(x')}) .$$

The right-hand side splits up into four parts, which we shall consider separately.

The first term is $|\tau|^2\mathbb{E}_{x,x'}\delta_1(x,x')$, which equals $|\tau|^2\|\delta_2\|_2^2$. The second is $\tau\mathbb{E}_{x,x'}\delta_1(x,x')\overline{v(x')}$. Now

$$\begin{aligned}\mathbb{E}_{x,x'}\delta_1(x,x')\overline{v(x')} &= \mathbb{E}_{x,x'}\mathbb{E}_y f(x,y)\overline{f(x',y)v(x')}\\ &= \mathbb{E}_y\delta_2(y)\mathbb{E}_{x'}\overline{f(x',y)v(x')}\\ &\leqslant \|\delta_2\|_2\Big(\mathbb{E}_y|\mathbb{E}_{x'}f(x',y)v(x')|^2\Big)^{1/2} ,\end{aligned}$$

which is at most $\|\delta_2\|_2 c_5^{1/2}\|v\|_2\|f\|_{U_2}$, by (v). Since both τ and $\|v\|_2$ are at most $\|u\|_2$ and

$$\|\delta_2\|_2^2=\mathbb{E}_{x,x'}\delta_1(x,x')\leqslant\big(\mathbb{E}_{x,x'}|\delta_1(x,x')|^2\big)^{1/2}=\|f\|_{U_2}^2 ,$$

it follows that the second term is at most $c_5^{1/2}\|f\|_{U_2}^2\|u\|_2^2$. The third term is just the complex conjugate of the second, so it is bounded in the same way. And finally the fourth term equals $\mathbb{E}_y|\mathbb{E}_x f(x,y)v(x)|^2$ and so by (v) it is at most $c_5\|f\|_{U_2}^2\|v\|_2^2\leqslant c_5\|f\|_{U_2}^2\|u\|_2^2$.

This shows that for any u we have the inequality

$$\mathbb{E}_{x,x'}\delta_1(x,x')u(x)\overline{u(x')}-|\mathbb{E}_x u(x)|^2\|\delta_2\|_2^2\leqslant(2c_5^{1/2}+c_5)\|f\|_{U_2}^2\|u\|_2^2 .$$

Now let us apply this inequality to the function $u(x) = \overline{f(x,y)}$, obtaining the inequality

$$\mathbb{E}_{x,x'}\delta_1(x,x')\overline{f(x,y)}f(x',y) - |\mathbb{E}_x f(x,y)|^2\|\delta_2\|_2^2 \leqslant (2c_5^{1/2}+c_5)\|f\|_{U_2}^2\mathbb{E}_x|f(x,y)|^2 \ .$$

Now let us take the expectation over y of both sides. The first term on the left-hand side has mean $\mathbb{E}_{x,x'}|\delta_1(x,x')|^2$, while the second equals $|\delta_2(y)|^2\|\delta_2\|_2^2$, which has mean $\|\delta_2\|_2^4$, which also equals $\Big(\mathbb{E}_{x,x'}\delta_1(x,x')\Big)^2$. Therefore, the expectation of the left-hand side is the variance of $\delta_1(x,x')$. The expectation of the right-hand side is $(2c_5^{1/2}+c_5)\|f\|_{U_2}^2\|f\|_2^2$, so we obtain (iii) with constant $c_3 = (2c_5^{1/2}+c_5)\|f\|_2^2/\|f\|_{U_2}^2$. □

We shall be particularly interested in the case where f is the characteristic function of a bipartite graph, and especially in the fact that property (iii) implies all the others. Notice that

$$\|\delta_2\|_2^2 = \mathbb{E}_y\big(\mathbb{E}_x f(x,y)\big)^2 = \mathbb{E}_{x,x'}\mathbb{E}_y f(x,y)f(x',y) = \mathbb{E}_{x,x'}\delta_1(x,x') \ ,$$

so (iii) is the statement that the variance of $\delta_1(x,x')$ is small compared with $\mathbb{E}_{x,x'}\delta_1(x,x')^2$. In the case of a bipartite graph, this says that almost all the intersections of pairs of neighbourhoods $N(x)$ and $N(x')$, with $x,x' \in X$, have about the same size.

This property should be compared with quasirandomness. If the average density of a neighbourhood is δ and almost all pairs of neighbourhoods intersect in a set of density approximately δ^2, then the graph is quasirandom. If, however, we have just the weaker property that the intersections of pairs of neighbourhoods have about the same size (without the further property that this size is almost minimal) then we obtain the weaker, but still useful, conclusion that G can be "quasirandomly approximated" by the very simple function $1 \otimes \delta_2$. In fact, the only difference between this and quasirandomness is that a quasirandom graph looks like a random graph where all edges are chosen independently with some probability p, whereas a graph with the weaker intersections property looks like a graph where the edges are again chosen independently, but now the probability of choosing the edge xy is given by some number $p(y) \in [0,1]$ that can depend on y. Furthermore, it is easy to identify $p(y)$: it is just the normalized degree of y.

We finish this section by remarking that the weak implication in Theorem 3.12 was necessarily so. To see this, suppose that f has singular-value decomposition $\sum_{i=1}^t \lambda_i u_i \otimes v_i$ and suppose that $u_1 = 1$. Then $ff^* = \sum_{i=1}^t |\lambda_i|^2 u_i \otimes u_i$, from which it follows that $\mathbb{E}_{x,x'}\delta_1(x,x')^2 = \sum_{i=1}^t |\lambda_i|^4$, as we already know. Now in this case $\delta_2(y) = \sum_{i=1}^t \lambda_i \mathbb{E}_x u_i(x) v_i(y)$, which equals $\lambda_1 v_1(y)$, since $\mathbb{E}_x u_1(x) = 1$ and $\mathbb{E}_x u_i(x) = 0$

for every $i > 1$ (because all the other u_i are orthogonal to the constant function 1). Therefore, $\|\delta_2\|_2^4 = |\lambda_1|^4$. It follows that the variance $\mathbb{E}_{x,x'}\Big|\delta_1(x,x') - \|\delta_2\|_2^2\Big|^2$ of $\delta_1(x,x')$ equals $\sum_{i>1}|\lambda_i|^4$. Notice that in this case property (iii) is exactly equivalent to properties (ii) and (iv).

It is an easy consequence of Lemma 3.4 that the norm of the restriction to functions of mean 0 of the linear map corresponding to f is equal to the largest value of $|\lambda_i|$ such that $i \neq 1$. Therefore, if we wish to bound the variance of $\delta_1(x,x')$ in terms of this norm, then we are trying to bound an ℓ_4-norm in terms of an ℓ_∞-norm, which we cannot do. However, we do at least have the inequality

$$\sum_{i\geqslant 2}|\lambda_i|^4 \leqslant \max_{i\geqslant 2}|\lambda_i|^2 \sum_{i=1}^{t}|\lambda_i|^2 = \max_{i\geqslant 2}|\lambda_i|^2\|f\|_2^2\ .$$

This is why it is not surprising that a ratio of the form $\|f\|_2^2/\|f\|_{U_2}^2$ appears in the proof that (v) implies (iii).

§4. Some weak regularity lemmas.

Theorem 3.12 would be just a curiosity were it not for the fact that graphs that satisfy property (iii) arise naturally in connection with a technique used in [G2] to prove a quantitative version of the Balog-Szemerédi theorem. In this section we shall begin by giving an answer to the following rather general-sounding question. Suppose you have a (not too sparse) bipartite graph G with vertex sets X and Y. How nice can you make G by restricting to a (not too small) subset $X' \subset X$? Notice that we are *not* allowed to restrict Y—for that question it is well known that one can restrict to a quasirandom graph. In our case, the best structure we could possibly hope for is exactly the structure described at the end of the previous section, namely a graph that resembles a random graph where the edge xy is chosen with probability $p(y)$. The reason for this is simple: if we actually *do* define a random graph this way, then with high probability, any restriction of the graph to a subset of the form $X' \times Y$ will still resemble a random graph where the edge xy is chosen with probability $p(y)$. The main result of this section is that we can actually find this structure. Once we have shown this (the proof is not very hard), we shall draw some consequences.

At the heart of the proof is the following lemma, which is a small generalization of Lemma 7.4 of [G2]. The slightly peculiar use of $(1+\eta)^{1/2}$ instead of $1+\eta$ is not an important feature of the statement: it just makes it a little bit more convenient to use later.

Lemma 4.1. *Let G be a bipartite graph with vertex sets X and Y of size m and n, and let γ, α, η and ϵ be positive constants. Suppose that there are at least ϵm^2 pairs $(x,x') \in X^2$ such that $|N(x) \cap N(x')| \geqslant \alpha(1+\eta)^{1/2}n$. Then there is a subset $B \subset X$ of size at least $(\epsilon\gamma)^{4\eta^{-1}\log(\alpha^{-1})}m$ such that $|N(x) \cap N(x')| \geqslant \alpha n$ for all but at most $\gamma|B|^2$ pairs $(x,x') \in B^2$.*

Proof. Let r be a positive integer to be chosen later. Pick $y_1,\dots,y_r$ in Y uniformly at random and let $B = N(y_1) \cap \dots \cap N(y_r)$. Define a pair (x,x') to be *good* if $|N(x) \cap N(x')| \geqslant \alpha(1+\eta)^{1/2}n$ and *bad* if $|N(x) \cap N(x')| < \alpha n$ and let the numbers of these pairs belonging to B be $g(B)$ and $b(B)$ respectively. A good pair will belong to B^2 with probability at least $\alpha^r(1+\eta)^{r/2}$, while a bad pair will belong with probability less than α^r, so the expected value of $g(B) - \gamma^{-1}b(B)$ is at least $\alpha^r m^2\Big(\epsilon(1+\eta)^{r/2} - \gamma^{-1}\Big)$. Let r be such that $\epsilon(1+\eta)^{r/2} - \gamma^{-1} \geqslant 1$. Then there must exist for this r a set B such that $g(B) - \gamma^{-1}b(B) \geqslant \alpha^r m^2$, which implies that $|B| \geqslant \alpha^{r/2}m$ and $b(B) \leqslant \gamma|B|^2$.

It remains to choose an appropriate value for r. A simple calculation shows that the inequality we needed is satisfied when $r = 8\eta^{-1}\log(\epsilon^{-1}\gamma^{-1})$ (the fact that this is not necessarily an integer is compensated for by the fact that the simple calculation was not optimized, so we shall ignore it). This gives us our set B with cardinality at least $\alpha^{4\eta^{-1}\log(\epsilon^{-1}\gamma^{-1})}m = (\epsilon\gamma)^{4\eta^{-1}\log(\alpha^{-1})}m$, which proves the lemma. $\square$

Corollary 4.2. *Let G be a bipartite graph with vertex sets X and Y of sizes m and n respectively, let δ, η, ϵ and γ be positive constants less than 1. Suppose that there are at least ϵm^2 pairs $(x,x') \in X^2$ such that $\delta_1(x,x') \geqslant \delta(1+\eta)^{1/2}$. Then there is a constant $\alpha \geqslant \delta$ and a subset $B \subset X$ of density at least $(\epsilon\gamma)^{8\eta^{-2}\log(\delta^{-1})^2}$ with the following two properties.*

(i) $\delta_1(x,x') \geqslant \alpha$ for all but at most $\gamma|B|^2$ pairs $(x,x') \in B^2$.

(ii) $\delta_1(x,x') \leqslant \alpha(1+\eta)$ for all but at most $\epsilon|B|^2$ pairs $(x,x') \in B^2$.

In particular, $\mathbb{E}_{x,x' \in B}|\delta_1(x,x') - \alpha|^2 \leqslant \epsilon + \alpha^2(\gamma + \eta^2)$.

Proof. We prove this result by iterating the previous lemma. Let $B_0 = X$ and let $\Delta_0 = \delta$. Since $\delta_1(x,x') \geqslant \Delta_0(1+\eta)^{1/2}$ for at least $\epsilon|B_0|^2$ pairs $(x,x') \in B_0^2$, Lemma 4.1 gives us a subset $B_1 \subset B_0$ of size at least $(\epsilon\gamma)^{4\eta^{-1}\log(\Delta_0^{-1})}|B_0|$ such that $\delta_1(x,x') \geqslant \Delta_0$ for all but at most $\gamma|B_1|^2$ pairs $(x,x') \in B_1^2$.

If we also have $\delta_1(x,x') \leqslant \Delta_0(1+\eta)$ for all but at most $\epsilon|B_1|^2$ pairs $(x,x') \in B_1^2$ then B_1 has the property we require of our set B. Otherwise, let $\Delta_1 = \Delta_0(1+\eta)^{1/2}$. We then know that $\delta_1(x,x') \geqslant \Delta_1(1+\eta)^{1/2}$ for at least $\epsilon|B_1|^2$ pairs $(x,x') \in B_1^2$, which is the assumption we first started with except that B_0 and Δ_0 have been replaced by B_1 and Δ_1.

We are therefore in a position to iterate, producing a sequence of sets $B_0 \supset B_1 \supset \dots$ and real numbers $\Delta_i = \delta(1+\eta)^{i/2}$. Since Δ_i cannot exceed 1, the iteration must stop after at most $t = 2\log_{1+\eta}(\delta^{-1}) \leqslant 2\eta^{-1}\log(\delta^{-1})$ steps, at which point we have found a set B with the required property. Moreover, each B_i has density at least $(\epsilon\gamma)^{4\eta^{-1}\log(\Delta_0^{-1})}$ times the density of B_{i-1}. Therefore the set B we eventually find has density at least $((\epsilon\gamma)^{4\eta^{-1}\log(\delta^{-1})})^t$, which is at least the bound stated.

The estimate for $\mathbb{E}_{x,x'\in B}|\delta_1(x,x') - \alpha|^2$ follows from the fact that if we choose a random pair $(x,x') \in B^2$, then the probability that $\delta_1(x,x') > \alpha(1+\eta)$ is at most ϵ, in which case we still know that $\delta_1(x,x') \leqslant 1$, and the probability that $\delta_1(x,x') < \alpha$ is at most γ, in which case we know that $\delta_1(x,x') \geqslant 0$. □

Remark. An interesting feature of the above result is that it is only the dependence on η that is exponentially expensive. Thus, it can be applied to quite sparse sets (not as sparse as $n^{99/100}$, say, but much sparser than $n/\log n$), and we can ask for ϵ to be very small. Indeed, on this second point, the method of proof is sufficiently flexible that if one is content with worse bounds, then one can ask for much stronger conclusions. Later in the paper we shall see that this can be useful.

Given a bipartite graph G with vertex sets X and Y, let us say that a subset $B \subset X$ is *(ϵ,γ,η)-regular for G with parameter α* if properties (i) and (ii) of Corollary 4.2 hold. The next result is a rather simple strengthening of Corollary 4.2, where the result is not a single (ϵ,γ,η)-regular set, but a partition of X into such sets. We shall use the word "density" in two senses: when it refers to a subset X' of X then it means $|X'|/|X|$ and when it refers to an induced bipartite subgraph $G(X',Y')$ of G then it means the number of edges between X' and Y' divided by $|X'||Y'|$.

Theorem 4.3. *Let δ, θ, ϵ and η be positive constants less than 1, suppose that $\epsilon < \delta$, and let G be a bipartite graph with vertex sets X and Y. Then there is a partition $X = B_0 \cup B_1 \cup \dots \cup B_r$ with the following properties.*

(i) For each $i \geqslant 1$ the set B_i has density at least $(\epsilon\gamma)^{8\eta^{-2}\log(\delta^{-1})^2}\theta$.

(ii) For each $i \geqslant 1$ the set B_i is (ϵ,γ,η)-regular for G.

(iii) For each $i \geqslant 1$ the graph $G(B_i,Y)$ has density at least $2\delta^{1/2}$.

(iv) For every subset $E \subset B_0$ of density at least θ, the graph $G(E,Y)$ has density less than $2\delta^{1/2}$.

Proof. If we cannot just set $X = B_0$ then there must be a subset $E \subset X$ of density at least θ such that the density of the induced subgraph $G(E,Y)$ is at least $2\delta^{1/2}$. Let us apply Corollary 4.2 to this subgraph. We know that $\mathbb{E}_{x,x'\in E}\delta_1(x,x') = \mathbb{E}_y\delta_E(y)^2 \geqslant (\mathbb{E}_y\delta_E(y))^2 \geqslant 4\delta$, so if (x,x') is a random pair chosen from E, then the probability that $\delta_1(x,x') \geqslant 2\delta$ is at least 2δ, and therefore the probability that $\delta_1(x,x') \geqslant \delta(1+\eta)^{1/2}$ is at least ϵ. Corollary 4.2 therefore gives us a set B_1 of density at least $(\epsilon\gamma)^{8\eta^{-2}\log(\delta^{-1})^2}\theta$ that is (ϵ,γ,η)-regular for G.

We now repeat. If we cannot set $B_0 = X \setminus B_1$, then there must be a subset $E \subset X \setminus B_1$ of density at least θ such that the density of the induced subgraph $G(E,Y)$ is at least $2\delta^{1/2}$. This gives us a set B_2 of density at least $(\epsilon\gamma)^{8\eta^{-2}\log(\delta^{-1})^2}\theta$ that is (ϵ,γ,η)-regular for G.

If we continue this iteration for as long as we can, then eventually we will find that the set $X \setminus (B_1 \cup \ldots \cup B_r)$ has no subset E of density θ such that the graph $G(E,Y)$ has density at least $2\delta^{1/2}$. We then set $B_0 = X \setminus (B_1 \cup \ldots \cup B_r)$ and the result is proved. $\square$

The above result can be thought of as a "weak regularity lemma", and it may help if we compare it with existing regularity lemmas. The most famous of these is Szemerédi's regularity lemma, which starts with a parameter ϵ and provides a partition of the vertices of a graph into sets $X_1,\ldots,X_k$, with k depending on ϵ only, in such a way that for at least $(1-\epsilon)k^2$ pairs (X_i,X_j) the induced bipartite subgraph $G(X_i,X_j)$ is ϵ-quasirandom. Although this is an extremely useful result, it has the drawback that the dependence of k on ϵ is very poor (and an example in [G1] shows that this is necessary).

Because of this, there are good reasons for trying to find alternative statements that are weak enough to give rise to better bounds, but still strong enough for applications. This turns out to be possible in some situations and tantalizingly difficult in others. One of the most useful weak regularity lemmas is due to Frieze and Kannan, and one way of describing how it relates to Szemerédi's lemma is as follows. A quasirandom bipartite graph of density p is a bipartite graph that is close in the U_2-norm to the constant function p. Suppose that $X_1,\ldots,X_k$ is the partition arising in Szemerédi's regularity lemma, and suppose that we define an averaging projection $PG(x,y)$ of G by setting $PG(x,y)$ to be the density of the bipartite graph $G(X_i,X_j)$ such that $x \in X_i$ and $y \in X_j$. Then Szemerédi's lemma says that G is well-approximated by PG in the U_2-norm on almost all pairs (X_i,X_j). The lemma of Frieze and Kannan provides a partition that gives rise to a more global notion of approximation: this time we cannot say anything about the U_2-norm of $G - PG$ if we restrict to a pair (X_i,X_j) but we can at least say that $\|G-PG\|_{U_2}$ is small. That is, G is quasirandomly approximated by PG in a global sense, but not necessarily if you restrict to the pairs (X_i, X_j).

What we have proved, though this will not become fully clear until after the next few results, is an "intermediate regularity lemma", where G is well-approximated by PG on sets of the form $X_i \times Y$. (This is a statement about bipartite graphs, though we shall obtain graph statements as well.) To summarize, for Szemerédi one can look at pairs of small sets, for Frieze and Kannan one can look at pairs of large sets, and here one can look at pairs of sets where one is small and the other large.

Theorem 4.3 provides us with a collection of graphs $G(B_i, Y)$ that are individually well-behaved. However, although they are edge disjoint, they all share a vertex set Y, and the next two lemmas can be used to show that they interact in an interesting way. Roughly speaking, we shall show that if $\delta_1(x, x')$ is approximately constant over pairs (x, x') from B_i^2 and also over pairs from B_j^2, then it is approximately constant over pairs from $B_i \times B_j$. One can make this assertion precise in different ways, according to whether one uses the stronger property (iii) or the weaker property (v) of Theorem 3.12 as the initial assumption. The next lemma uses property (iii) and the one after it uses property (v). In each case, the conclusion resembles the premise in an obvious way.

We shall state our results in terms of two sets X and X'. Variables such as x, x_1 and x_2 will always be assumed to range over X, while variables such as x', x_1' and x_2' will range over X'. Given two functions $f : X \times Y \to \mathbb{C}$ and $f' : X' \times Y \to \mathbb{C}$, we shall write $\delta_2(y)$ for $\mathbb{E}_x f(x, y)$ and $\delta_2'(y)$ for $\mathbb{E}_{x'} f'(x', y)$. The meaning of δ_1 will depend in an obvious way on the variables inside it: for example, $\delta_1(x, x') = \mathbb{E}_y f(x, y)\overline{f'(x', y)}$.

Lemma 4.4. *Let X, X' and Y be finite sets, let $f : X \times Y \to \mathbb{C}$ and let $f' : X' \times Y \to \mathbb{C}$. Let $0 < c \leqslant 2^{-24}$, and suppose that $\mathbb{E}_{x_1,x_2}\Big|\delta_1(x_1, x_2) - \|\delta_2\|_2^2\Big|^2 \leqslant c\|f\|_{U_2}^4$ and that $\mathbb{E}_{x_1',x_2'}\Big|\delta_1(x_1', x_2') - \|\delta_2'\|_2^2\Big|^2 \leqslant c\|f'\|_{U_2}^4$. Then $\mathbb{E}_{x,x'}\Big|\delta_1(x, x') - \langle\delta_1, \delta_2\rangle\Big|^2 \leqslant 16c^{1/16}\|f\|_{U_2}^2\|f'\|_{U_2}^2$.*

Proof. By Theorem 3.12, we know that there is a constant c_2, with a power-type dependence on c, such that $\|f - 1 \otimes \delta_2\|_{U_2}^4 \leqslant c_2\|f\|_{U_2}^4$ and $\|f' - 1 \otimes \delta_2'\|_{U_2}^4 \leqslant c_2\|f'\|_{U_2}^4$. We also know that

$$\begin{aligned}\delta_1(x, x') - \langle\delta_1, \delta_2\rangle &= \mathbb{E}_y\big(f(x, y)\overline{f'(x', y)} - \delta_2(y)\overline{\delta_2'(y)}\big) \\ &= \mathbb{E}_y(f(x, y) - \delta_2(y))\overline{f'(x', y)} + \mathbb{E}_y\delta_2(y)(\overline{f'(x', y)} - \overline{\delta_2'(y)}) \\ &= \mathbb{E}_y g(x, y)\overline{f'(x, y)} + \mathbb{E}_y\delta_2(y)\overline{g'(x, y)}\ ,\end{aligned}$$

where we have set $g = f - 1 \otimes \delta_2$ and $g' = f' - 1 \otimes \delta_2'$. By Minkowski's inequality, we may therefore deduce that $\left(\mathbb{E}_{x,x'}\left|\delta_1(x,x') - \langle\delta_1,\delta_2\rangle\right|^2\right)^{1/2}$ is at most

$$\left(\mathbb{E}_{x,x'}\left|\mathbb{E}_y g(x,y)\overline{f'(x',y)}\right|^2\right)^{1/2} + \left(\mathbb{E}_{x,x'}\left|\mathbb{E}_y \delta_2(y)\overline{g'(x',y)}\right|^2\right)^{1/2} .$$

Let us estimate these two terms separately. First,

$$\begin{aligned}
&\mathbb{E}_{x,x'}\left|\mathbb{E}_y g(x,y)\overline{f'(x',y)}\right|^2 \\
&\qquad = \mathbb{E}_{x,x'}\mathbb{E}_{y,y'} g(x,y)\overline{g(x,y')f'(x',y)}f'(x',y) \\
&\qquad = \mathbb{E}_{y,y'}\left(\mathbb{E}_x g(x,y)\overline{g(x,y')}\right)\left(\mathbb{E}_{x'}\overline{f'(x',y)}f'(x',y)\right) \\
&\qquad \leqslant \left(\mathbb{E}_{y,y'}\left|\mathbb{E}_x g(x,y)\overline{g(x,y')}\right|^2\right)^{1/2}\left(\mathbb{E}_{y,y'}\left|\mathbb{E}_{x'}\overline{f'(x',y)}f'(x',y)\right|^2\right)^{1/2} \\
&\qquad = \|g\|_{U_2}^2\|f'\|_{U_2}^2 .
\end{aligned}$$

Similarly, the square of the second term is at most $\|1\otimes\delta_2\|_{U_2}^2\|g'\|_{U_2}^2 = \|\delta_2\|_2^2\|g'\|_{U_2}^2 \leqslant \|f\|_{U_2}^2\|g'\|_{U_2}^2$.

It follows that the sum of the two terms is at most $2c_2^{1/4}\|f\|_{U_2}\|f'\|_{U_2}$, so we may take $c' = 4c_2^{1/2}$. Looking carefully at the bounds in Theorem 3.12, one can check that this is at most $16c^{1/16}$ when $c \leqslant 2^{-24}$. □

Lemma 4.4 tells us that if the variance of δ_1 is small when it is restricted to X^2 and small when it is restricted to X'^2, then the variance is small when it is restricted to $X \times X'$. The next lemma uses property (v) of Theorem 3.12 and arrives at a slightly weaker and more global conclusion, which implies in particular that, however you choose subsets $A \subset X'$ and $A' \subset X'$ that are not too small, the average $\mathbb{E}_{x\in A}\mathbb{E}_{x'\in A'}\delta_1(x,x')$ will be roughly the same.

Lemma 4.5. *Let X, X' and Y be finite sets, let $f : X \times Y \to \mathbb{C}$ and let $f' : X' \times Y \to \mathbb{C}$. Suppose that $\mathbb{E}_y|\mathbb{E}_x f(x,y)v(x)|^2 \leqslant c\|f\|_{U_2}^2\|v\|_2^2$ whenever $\mathbb{E}_x v(x) = 0$, and $\mathbb{E}_y|\mathbb{E}_{x'} f'(x',y)v'(x)|^2 \leqslant c\|f'\|_{U_2}^2\|v'\|_2^2$ whenever $\mathbb{E}_{x'}v'(x') = 0$. Then*

$$\left|\mathbb{E}_{x,x'}\delta_1(x,x')u(x)u'(x') - \mathbb{E}_{x,x'}u(x)u'(x')\langle\delta_2,\delta_2'\rangle\right| \leqslant c^{1/2}\|f\|_{U_2}\|f'\|_{U_2}\|u\|_2\|u'\|_2 .$$

Proof. Let $u : X \to \mathbb{C}$ and $u' : X' \to \mathbb{C}$ be arbitrary functions. Let $\tau = \mathbb{E}_x u(x)$ and $\tau' = \mathbb{E}_{x'}u(x')$, and let $u(x) = \tau + v(x)$ and $u'(x') = \tau' + v'(x')$. Let us write F and F' for the linear maps from X and X' to Y associated with the functions f and f'. That is, $Fu(y) = \mathbb{E}_x f(x,y)u(x)$ and $F'u'(y) = \mathbb{E}_{x'} f'(x',y)u'(x')$. We shall also write τ and τ' for the constant functions that take values τ and τ'.

Note first that, for any $u : X \to \mathbb{C}$,

$$\|Fu\|_2^2 = \mathbb{E}_{x_1,x_2}\delta_1(x_1,x_2)u(x_1)\overline{u(x_2)} \leqslant \Big(\mathbb{E}_{x_1,x_2}\delta_1(x_1,x_2)^2\Big)^{1/2}\|u\|_2^2 = \|f\|_{U_2}^2\|u\|_2^2 \ ,$$

so the norm of F before we restrict it to the space of functions of mean zero is at most $\|f\|_{U_2}$. (This also follows easily from Lemma 3.4.) Similarly, we have the corresponding result for F' and f'. It follows that

$$\begin{aligned}|\langle Fu, F'u'\rangle - \langle F\tau, F'\tau'\rangle| &\leqslant |\langle Fv, F'\tau'\rangle| + |\langle F\tau, F'v'\rangle| \\ &\leqslant \|Fv\|_2\|F'\tau'\|_2 + \|F\tau\|_2\|F'v'\|_2 \\ &\leqslant c^{1/2}\|f\|_{U_2}\|f'\|_{U_2}(\|v\|_2\|\tau'\|_2 + \|\tau\|_2\|v'\|_2 \\ &\leqslant c^{1/2}\|f\|_{U_2}\|f'\|_{U_2}\|u\|_2\|u'\|_2 \ .\end{aligned}$$

Now $F\tau = \tau\delta_2$, $F'\tau' = \tau'\delta_2'$, and $\langle Fu, F'u'\rangle = \mathbb{E}_{x,x'}\delta_1(x,x')u(x)u'(x')$, so this implies the result. □

Of particular interest to us is the obvious graph-theoretical consequence of the above facts: if f, f', u and u' all take values in $\{0,1\}$, then f and f' are the adjacency matrices of bipartite graphs $G \subset X \times Y$ and $G' \subset X' \times Y$, and u and u' are characteristic functions of subsets $A \subset X$ and $A' \subset X'$. Then $\langle Fu, F'u'\rangle$ is the number of paths of length 2 from A to A', and Lemmas 4.4 and 4.5 imply that this is close to $|A||A'|\langle\delta_2,\delta_2'\rangle$. That is, the product of G and G' has roughly the same density $\langle\delta_2,\delta_2'\rangle$ everywhere.

Theorem 4.3 therefore gives us a strong regularity statement for "squares" of graphs, or more accurately for the function $\delta_1(x,x')$ (which corresponds to the product of the adjacency matrix of G with its transpose). It tells us that we can partition the vertex set of G into sets $B_0, B_1, \dots, B_r$ in such a way that there are almost no edges out of B_0 and such that for *every* $i,j \geqslant 1$ there exists α_{ij} such that $\delta_1(x,x')$ is approximately α_{ij} for almost every $(x,x') \in B_i \times B_j$. This is a significantly stronger statement than can be made about the partition in the weak regularity lemma of Frieze and Kannan, but the bound is of comparable size.

We now make this statement precise, beginning with a technical lemma.

Lemma 4.6. *Let $0 < \delta < 1$ and $0 < c < 1/100$, and let $\epsilon = \delta^2 c/4$ and $\eta = c^{1/2}/2$. Let G be a bipartite graph of density δ with vertex sets X and Y and suppose that $B \subset X$ is (ϵ,ϵ,η)-regular for G, with a parameter $\alpha \geqslant \delta$, and let $\delta_B(y)$ denote the normalized degree in B of a vertex $y \in Y$. Then*

$$\mathbb{E}_{x,x'\in B}\Big|\delta_1(x,x') - \|\delta_B\|^2\Big|^2 \leqslant c\mathbb{E}_{x,x'\in B}\delta_1(x,x')^2 \ .$$

Proof. By the last part of the conclusion of Corollary 4.2,

$$\mathbb{E}_{x,x'\in B}|\delta_1(x,x')-\alpha|^2 \leqslant \delta^2 c/2 + \alpha^2 c/4 \leqslant 3\alpha^2 c/4 \ .$$

It follows that $\Big(\mathbb{E}_{x,x'\in B}|\delta_1(x,x')|^2\Big)^{1/2} \geqslant \alpha - (3\alpha^2 c/4)^{1/2} \geqslant \alpha(1-\sqrt{3c}/2)$, so we have shown that

$$\mathbb{E}_{x,x'\in B}|\delta_1(x,x')-\alpha|^2 \leqslant \frac{3c}{4(1-\sqrt{3c}/2)^2}\mathbb{E}_{x,x'\in B}\delta_1(x,x')^2 \ .$$

It can be checked that this is at most $c\mathbb{E}_{x,x'\in B}\delta_1(x,x')^2$ when $c < 1/100$. The result follows, since $\|\delta_{B_i}\|_2^2$ is the mean of δ_1 over B^2. □

Theorem 4.7. *Let $\zeta, \theta > 0$ and let G be a bipartite graph with vertex sets X and Y. Then there is a partition of X into sets $B_0, B_1, \dots, B_r$, with $r \leqslant \theta^{-1}\exp(L\zeta^{-16}(\log(\zeta^{-1})^3)$ for some absolute constant L, with the following property. For each i and each $y \in Y$ let $\delta_{B_i}(y)$ be the normalized degree of y in B_i. Then*

$$\mathbb{E}_{x\in B_i}\mathbb{E}_{x'\in B_j}\Big|\delta_1(x,x') - \langle\delta_{B_i},\delta_{B_j}\rangle\Big|^2 \leqslant \zeta$$

whenever $i, j \geqslant 1$, and either B_0 has density less than θ or the same inequality holds for $i, j \geqslant 0$.

Proof. Let $c = 2^{-64}\zeta^{16}$, $\delta = \zeta^2/4$, $\epsilon = \gamma = \delta^2 c/4$ and $\eta = c^{1/2}/2$, and let $B_0, B_1, \dots, B_r$ be the partition arising in Theorem 4.3 for these parameters.

For each i let f_i denote the restriction of the characteristic function of G to the set $B_i \times Y$. If $i \geqslant 1$, then B_i is (ϵ,γ,η)-regular for G and has density at least $2\delta^{1/2}$. Therefore, by Lemma 4.6,

$$\mathbb{E}_{x,x'\in B_i}|\delta_1(x,x') - \|\delta_{B_i}|^2 \leqslant c\mathbb{E}_{x,x'\in B_i}\delta_1(x,x')^2 = \|f_i\|_{U_2}^4 \ .$$

We now have the conditions of Lemma 4.4 for any pair of sets B_i and B_j with $i, j \geqslant 1$. It follows that

$$\mathbb{E}_{x\in B_i}\mathbb{E}_{x'\in B_j}|\delta_1(x,x') - \langle\delta_{B_i},\delta_{B_j}\rangle|^2 \leqslant \zeta\|f_i\|_{U_2}^2\|f_j\|_{U_2}^2$$

for every such pair.

We also know that

$$\mathbb{E}_{x\in B_0}\mathbb{E}_{x'\in B_i}|\delta_1(x,x')|^2 \leqslant \|f_0\|_{U_2}^2\|f_i\|_{U_2}^2 \ ,$$

by Lemma 2.2. But either B_0 has density at most θ, in which case we are done, or the graph $G(B_0, Y)$ has density at most $2\delta^{1/2}$, from which it follows that $\|f_0\|_{U_2}^4 \leqslant \|f_0\|_2^4 \leqslant \|f_0\|_2^2 \|f_0\|_\infty^2 \leqslant 4\delta$. (The first of these inequalities can be deduced easily from Lemma 2.4, or else it is an easy exercise to prove it directly.) Since the mean of $\delta_1(x, x')$ over $B_0 \times B_i$ is $\langle \delta_{B_0}, \delta_{B_i} \rangle$, it follows that

$$\mathbb{E}_{x \in B_0} \mathbb{E}_{x' \in B_i} |\delta_1(x, x') - \langle \delta_{B_0}, \delta_{B_i} \rangle|^2 \leqslant 2\delta^{1/2} \|f_i\|_{U_2}^2 \leqslant \zeta \ .$$

This shows that the partition has the desired properties, and the bounds of Theorem 4.3 can be used to check the upper bound on the size of r. □

Remark. Essentially the same proof gives a non-symmetric version of the result, in which one obtains a strong regularity lemma for general graph products. That is, given a bipartite graph G with vertex sets X and Y, and a bipartite graph G' with vertex sets X' and Y, one can find partitions of X and X' such that for almost every B in the first partition and B' in the second partition the function $\delta_1(x, x')$ (defined to be the density of the intersection of the G-neighbourhood of x with the G'-neighbourhood of x') is roughly constant over $B \times B'$. In the case $X = X'$ one can even ask for the two partitions to be the same, by proving a version of Lemma 4.1 that works for more than one graph at the same time. (This one can do, for example, by considering a bipartite graph from X to Y^2 that joins x to (y, y') if and only if $(x, y) \in G$ and $(x, y') \in G'$.)

We now prove a strengthening of Lemma 4.4 that allows us to exploit more fully the fact that we can obtain (ϵ, η) regular sets with ϵ much smaller than η. The next lemma shows that if B is (ϵ, η)-regular for a graph G, then all subsets $C \subset B$ have images, under the linear map given by the adjacency matrix of the graph $G(B, Y)$, that are approximately proportional to each other, provided that they are not too small—where the lower bound we require on the size tends to zero with ϵ and so can be chosen to be very small. The one after it will show that if B_i and B_j are (ϵ, η)-regular with a very small ϵ, then $\delta_1(x, x')$ is approximately equal to α_{ij} for almost every $(x, x') \in C_i \times C_j$, when C_i and C_j are subsets of B_i and B_j that are not too small—again with a lower bound that can be chosen to be very small.

Lemma 4.8. *Let δ, η, ϵ and κ be positive constants such that $\eta \leqslant 1/4$, $\kappa \leqslant 1$ and $\epsilon\kappa^{-2} \leqslant \delta\eta$. Let X and Y be finite sets and let G be a bipartite graph with vertex sets X and Y. Suppose that $B \subset X$ is $(\epsilon, \epsilon, \eta)$-regular for G with an α that is at least as big as δ. Let C be a subset of B of size at least $\kappa|B|$. For each $y \in Y$, let $\delta_B(y)$ and $\delta_C(y)$ be* *the normalized degrees of y in the induced subgraphs $G(B,Y)$ and $G(C,Y)$, respectively. Then $\|\delta_B-\delta_C\|_2^2\leqslant 8\eta\|\delta_B\|_2^2$.*

Proof. By assumption, there exists $\alpha\geqslant\delta$ such that $\delta_1(b_1,b_2)$ lies between α and $\alpha(1+\eta)$ for all but at most $\epsilon|B|^2$ pairs $(b_1,b_2)\in B^2$. Therefore, if $|C|=\kappa|B|$, then $\delta_1(b,c)$ lies between α and $\alpha(1+\eta)$ for all but at most $\epsilon\kappa^{-1}|B||C|$ pairs $(b,c)\in B\times C$. It follows that

$$\langle\delta_B,\delta_C\rangle=\mathbb{E}_y\delta_B(y)\delta_C(y)=\mathbb{E}_{c\in C}\mathbb{E}_{b\in B}\delta_1(b,c)$$

lies between $\alpha-\epsilon\kappa^{-1}$ and $\alpha(1+\eta)+\epsilon\kappa^{-1}$.

Similar arguments show that $\|\delta_B\|_2^2$ lies between $\alpha-\epsilon$ and $\alpha(1+\eta)+\epsilon$ and that $\|\delta_C\|_2^2$ lies between $\alpha-\epsilon\kappa^{-2}$ and $\alpha(1+\eta)+\epsilon\kappa^{-2}$. Since $\epsilon\kappa^{-2}\leqslant\alpha\eta$, all three of $\|\delta_B\|_2^2$, $\|\delta_C\|_2^2$ and $\langle\delta_B,\delta_C\rangle$ lie between $\alpha(1-\eta)$ and $\alpha(1+2\eta)$. It follows that

$$\|\delta_B-\delta_C\|_2^2\leqslant 2\alpha(1+2\eta)-2\alpha(1-\eta)=6\alpha\eta\leqslant 8\eta\|\delta_B\|_2^2\ .\qquad\square$$

Lemma 4.9. *Let δ, η, ϵ and κ be positive constants such that $\eta\leqslant 1/8$, $\kappa\leqslant 1$ and $\epsilon\kappa^{-2}\leqslant\delta^2\eta^2$. Let X and Y be finite sets and let G be a bipartite graph with vertex sets X and Y. Suppose that B and B' are subsets of X that are both (ϵ,ϵ,η)-regular for G with parameters α and α', respectively, that are at least as big as δ. Let C and C' be subsets of B and B' with sizes at least $\kappa|B|$ and $\kappa|B'|$, respectively. Let f_B, f'_B, f_C and $f_{C'}$ denote the restrictions of the characteristic function of G to the sets $B\times Y$, $B'\times Y$, $C\times Y$ and $C'\times Y$, respectively, and for each $y\in Y$, let $\delta_B(y)$, $\delta_C(y)$, $\delta_{B'}(y)$ and $\delta_{C'}(y)$ be the normalized degrees of y in the corresponding induced subgraphs. Then there is an absolute constant L such that*

$$\mathbb{E}_{x\in C}\mathbb{E}_{x'\in C'}|\delta_1(x,x')-\langle\delta_B,\delta_{B'}\rangle|^2\leqslant L\eta^{1/8}\|f_B\|_{U_2}^2\|f_{B'}\|_{U_2}^2\ .$$

Proof. We begin by proving a similar inequality with $\langle\delta_C,\delta_{C'}\rangle$ instead of $\langle\delta_B,\delta_{B'}\rangle$. We then deduce from Lemma 4.8 that the two inner products are approximately equal. The basic argument is simple but there are some tedious details involved in passing from an estimate in terms of one quantity to a similar estimate in terms of another that is approximately equal to it.

To prove the inequality with $\langle\delta_C,\delta_{C'}\rangle$ we use Lemma 4.4. Our assumptions tell us that $\delta_1(x_1,x_2)$ lies between α and $\alpha(1+\eta)$ for all but at most $\epsilon|B|^2$ pairs $(x_1,x_2)\in B^2$, and hence for all but at most $\epsilon\kappa^{-2}|C|^2$ pairs $(x_1,x_2)\in C^2$. Therefore, the variance of $\delta_1(x_1,x_2)$ over C^2 is at most $\epsilon\kappa^{-2}+\eta^2\alpha^2$, which is at most $2\eta^2\alpha^2$. Similarly, the variance of $\delta_1(x'_1,x'_2)$ over C'^2 is at most $2\eta^2\alpha^2$.

Let us write f_C for the restriction of f to $C \times Y$. Then the mean of $\delta_1(x_1, x_2)$ over C^2 is $\|\delta_C\|_2^2$, while the mean of $\delta_1(x_1, x_2)^2$ is $\|f_C\|_{U_2}^4$. Therefore, $\|\delta_C\|_2^2 \leqslant \|f_C\|_{U_2}^2$. From the proof of Lemma 4.8, we know that $\|\delta_C\|_2^2$ is at least $\alpha(1-\eta)$. (There we assumed a different bound for κ, but the assumption was weaker so the conclusion is valid here too.) But $(1-\eta)^2 > 1/2$, so $\|f_C\|_{U_2}^4 \geqslant \|\delta_C\|_2^4 \geqslant \alpha^2/2$. Therefore, our statement about the variance of $\delta_1(x_1, x_2)$ over C^2 implies that

$$\mathbb{E}_{x_1, x_2 \in C} \left| \delta_1(x_1, x_2) - \|\delta_C\|_2^2 \right|^2 \leqslant 4\eta^2 \|f_C\|_{U_2}^4 \ .$$

The same argument shows that

$$\mathbb{E}_{x_1', x_2' \in C'} \left| \delta_1(x_1', x_2') - \|\delta_{C'}\|_2^2 \right|^2 \leqslant 4\eta^2 \|f_{C'}\|_{U_2}^4 \ .$$

We therefore have the hypotheses of Lemma 4.4, with C and C' replacing X and X', with f_C and $f_{C'}$ replacing f and f', and with $c = 4\eta^2$. It follows from that lemma that

$$\mathbb{E}_{x \in C} \mathbb{E}_{x' \in C'} \left| \delta_1(x, x') - \langle \delta_C, \delta_{C'} \rangle \right|^2 \leqslant K\eta^{1/8} \|f_C\|_{U_2}^2 \|f_{C'}\|_{U_2}^2$$

for some absolute constant K.

It remains to recast this conclusion slightly. By Lemma 4.8 we know that $\|\delta_B - \delta_C\|_2^2 \leqslant 8\eta \|\delta_B\|_2^2$ and $\|\delta_{B'} - \delta_{C'}\|_2^2 \leqslant 8\eta \|\delta_{B'}\|_2^2$. Therefore,

$$\begin{aligned} |\langle \delta_B, \delta_{B'} \rangle - \langle \delta_C, \delta_{C'} \rangle| &\leqslant |\langle \delta_B, \delta_{B'} - \delta_{C'} \rangle| + |\langle \delta_B - \delta_C, \delta_{C'} \rangle| \\ &\leqslant 2(8\eta)^{1/2} \|\delta_B\|_2 \|\delta_{B'}\|_2 \ . \end{aligned}$$

Since $\langle \delta_C, \delta_{C'} \rangle$ is the mean of $\delta_1(x, x')$ over $C \times C'$, we have the equality

$$\begin{aligned} &\mathbb{E}_{x \in C} \mathbb{E}_{x' \in C'} \left| \delta_1(x, x') - \langle \delta_B, \delta_{B'} \rangle \right|^2 \\ &\qquad = \mathbb{E}_{x \in C} \mathbb{E}_{x' \in C'} \left| \delta_1(x, x') - \langle \delta_C, \delta_{C'} \rangle \right|^2 + \left| \langle \delta_B, \delta_{B'} \rangle - \langle \delta_C, \delta_{C'} \rangle \right|^2 \ , \end{aligned}$$

which, by the estimates we have obtained, is at most $K\eta^{1/8} \|f_C\|_{U_2}^2 \|f_{C'}\|_{U_2}^2 + 32\eta \|\delta_B\|_2^2 \|\delta_{B'}\|_2^2$.

Finally, we note that $\|f_C\|_{U_2}^4 = \mathbb{E}_{x_1, x_2 \in C} \delta_1(x_1, x_2)^2 \leqslant \|\delta_C\|^4 + 4\eta^2 \|f_C\|_{U_2}^4$, by our inequality for the variance earlier. Since $\eta < 1/8$, it follows that $\|f_C\|_{U_2}^4 \leqslant 2\|\delta_C\|_2^4$. Again because $\eta < 1/8$ we know from the fact that $\|\delta_B - \delta_C\|_2^2 \leqslant 8\eta \|\delta_B\|_2^2$ that $\|\delta_C\|_2 \leqslant 2\|\delta_B\|_2$. And we also know that $\|\delta_B\|_2 \leqslant \|f_B\|_{U_2}$. Putting this together, we find that $\|f_C\|_{U_2}^4 \leqslant 32\|f_B\|_{U_2}^4$. Similarly, $\|f_{C'}\|_{U_2}^4 \leqslant 32\|f_{B'}\|_{U_2}^4$. Therefore,

$$\mathbb{E}_{x \in C} \mathbb{E}_{x' \in C'} \left| \delta_1(x, x') - \langle \delta_B, \delta_{B'} \rangle \right|^2 \leqslant (32\eta + 32K\eta^{1/8}) \|f_B\|_{U_2}^2 \|f_{B'}\|_{U_2}^2 \ ,$$

which proves the lemma. □

§5. An application to additive combinatorics.

Let N be a positive integer, let $\mathbb{Z}_N$ denote the set of integers mod N and let A be a subset of $\mathbb{Z}_N$ of size $\lfloor N/2 \rfloor$. If $x \in \mathbb{Z}_N$ is chosen randomly, then the expected size of $A \cap (A + x)$ is $|A|^2/N$, which is approximately $N/4$. Must there exist some x such that $A \cap (A + x)$ has approximately this size?

In general, the answer is obviously no: if N is even, then A could be the set of all even elements of $\mathbb{Z}_N$, and then $A \cap (A+x)$ would have size 0 or $N/2$ for every x. However, there is no such easy counterexample if N is prime, so what happens in that case?

This question has an odd history. I thought of it while working on another problem, then thought I had a solution, and then realized that my solution was incorrect (because a certain error term was larger than a certain main term). But I liked the question, so I put it on a sheet of questions to accompany a course I gave on additive combinatorics. However, it became clear that it was more than just an exercise: none of the students attending the course managed to solve it, and it was some time before a solution was finally obtained, by Green and Konyagin [GK]. The best bound now known is due to Sanders [S], who proved that there is some x such that $\Big||A \cap (A + x)| - N/4\Big|$ is at most $CN(\log N)^{-1/3}$.

In this section, we shall give a new proof of this result. We obtain a considerably worse bound, so the interest is in the method rather than the conclusion (though it is not out of the question that one might be able to obtain better bounds by developing this method). The novel feature of the method is that it is "purely combinatorial", in the sense that it uses no Fourier analysis. That is, all our arguments take place in "physical space", whereas Fourier analysis was essential to the arguments of Green and Konyagin and of Sanders.

The first step is to prove a variant of Corollary 4.2. As mentioned earlier, it can be useful to strengthen the conclusion of this result, and that is what we shall do now. The extra strength is that the number of pairs for which $\delta_1(x, x') > \alpha(1+\eta)$ is not just a small fraction of $|B|^2$, but a small fraction that depends on the density of B.

Lemma 5.1. *Let G be a bipartite graph with vertex sets X and Y of sizes m and n respectively, let δ, η, ϵ and γ be positive constants less than 1. Suppose that there are at least ϵm^2 pairs $(x, x') \in X^2$ such that $\delta_1(x, x') \geqslant \delta(1 + \eta)^{1/2}$. Then there is a constant $\alpha \geqslant \delta$ and a subset $B \subset X$ with the following properties.*

(i) B has density $\beta \geqslant \exp(-T)$, where $T = (8\eta^{-1}\log\delta^{-1})^{2\eta^{-1}\log\delta^{-1}}(\log\epsilon^{-1}+\log\gamma^{-1})$. In particular, if ϵ and γ are both at least $\eta^2\delta^2$ and we set $t = 4\eta^{-1}\log\delta^{-1}$, then the density of B is at least $\exp(-t^t)$.

(ii) B is $(\epsilon\beta^2, \gamma, \eta)$-regular with parameter α.

Proof. The argument is very similar to the proof of Corollary 4.2. Let $B_0 = X$, let $\beta_0 = 1$ and let $\Delta_0 = \delta$. Since $\delta_1(x, x') \geqslant \Delta_0(1+\eta)^{1/2}$ for at least $\epsilon\beta_0^2|B_0|^2$ pairs $(x, x') \in B_0^2$, Lemma 4.1 gives us a subset $B_1 \subset B_0$ of size at least $(\epsilon\beta_0^2\gamma)^{4\eta^{-1}\log(\Delta_0^{-1})}|B_0|$ such that $\delta_1(x, x') \geqslant \Delta_0$ for all but at most $\gamma|B_1|^2$ pairs $(x, x') \in B_1^2$. Let β_1 be the density of B_1.

If we also have $\delta_1(x, x') \leqslant \Delta_0(1+\eta)$ for all but at most $\epsilon\beta_1^2|B_1|^2$ pairs $(x, x') \in B_1^2$ then B_1 has the property we require of our set B. Otherwise, let $\Delta_1 = \Delta_0(1+\eta)^{1/2}$. We then know that $\delta_1(x, x') \geqslant \Delta_1(1+\eta)^{1/2}$ for at least $\epsilon\beta_1^2|B_1|^2$ pairs $(x, x') \in B_1^2$, which is the assumption we first started with except that B_0, β_0 and Δ_0 have been replaced by B_1, β_1 and Δ_1.

We are therefore in a position to iterate. Again the iteration must stop after at most $t = 2\log_{1+\eta}\delta^{-1} \leqslant 2\eta^{-1}\log\delta^{-1}$ steps. However, this time what we can say about the density of the B_i is weaker: we know that $\beta_i \geqslant (\epsilon\beta_{i-1}^2\gamma)^{4\eta^{-1}\log(\delta^{-1})}$. Taking logarithms and writing $a_i = \log\beta_i^{-1}$, we have $a_0 = 0$ and

$$a_i \leqslant 4\eta^{-1}\log\delta^{-1}(\log\epsilon^{-1} + \log\gamma^{-1} + 2a_{i-1}) ,$$

from which it is an easy exercise to deduce that $a_i \leqslant (8\eta^{-1}\log\delta^{-1})^i(\log\epsilon^{-1} + \log\gamma^{-1})$. This implies the lower bound stated for the density of B. □

The next lemma shows that we can pass to a large subset of B with an even stronger property. There are some complicated relationships between the parameters involved, but the point to bear in mind is that in applications we would like κ_0 to be very small. Since it does not cost too much to make ϵ very small, we can have this, even if γ and η are fixed.

Lemma 5.2. *Let $\delta, \epsilon > 0$ and let $0 < \eta \leqslant 1/4$ and $0 < \gamma < \eta/4$. Let G be a bipartite graph with vertex sets X and Y and suppose that $B_0 \subset X$ is (ϵ, γ, η)-regular for G with $\alpha \geqslant \delta$. Let $\kappa_0 = (\epsilon/\delta\eta)^{1/2}$ and suppose that $\kappa_0 \leqslant \eta/2$. Then B_0 has a subset B of cardinality at least $|B_0|/2$ such that $\|\delta_C - \delta_B\|_2^2 \leqslant 24\eta\|\delta_B\|_2^2$ for every subset $C \subset B$ of size at least $\kappa_0|B_0|$.*

Proof. Let B be any subset of B_0 of size at least $|B_0|/2$, let $C \subset B$ be a subset of size $\kappa|B|$, with $\kappa \geqslant \kappa_0$, let $\theta = 24\eta$ and suppose that $\|\delta_C - \delta_B\|_2^2 \geqslant \theta\|\delta_B\|_2^2$.

The main fact on which our calculations will depend is that for any subset $D \subset B$ with $|D| \geqslant \kappa_0|B_0|$ we have $\|\delta_D\|_2^2 \leqslant \alpha(1+2\eta) \leqslant (1+4\eta)\|\delta_B\|_2^2$. To see the first inequality, we use the fact that there are at most $\epsilon|B_0|^2$ pairs $(x,x') \in B^2$ such that $\delta_1(x,x') \geqslant \alpha(1+\eta)$, from which it follows that

$$\|\delta_D\|_2^2 = \mathbb{E}_{x,x' \in D}\delta_1(x,x') \leqslant \kappa_0^{-2}\epsilon + \alpha(1+\eta) \leqslant \alpha(1+2\eta) \ ,$$

as claimed. As for the second, we know that $\|\delta_B\|_2^2 \geqslant \alpha(1-4\gamma) \geqslant \alpha(1-\eta)$. Since $\eta \leqslant 1/4$, one can check that $1+2\eta \leqslant (1+4\eta)(1-\eta)$.

An annoying technicality is that we have to show independently that κ cannot be greater than $1-\kappa_0$. Let us therefore suppose that $C \subset B$ has size $\kappa|B|$ with $\kappa > 1-\kappa_0$ and prove that $\|\delta_C - \delta_B\|_2 < \theta\|\delta_B\|_2$. We shall use repeatedly the fact that if D is any subset of B (or indeed of X) then $|D|\delta_D$ is the vector $\sum_{x\in D} N_x$, where N_x is the characteristic function of the neighbourhood of x. We shall also use the fact that the vectors N_x are positive, so that if $D \subset E$ then the norm of $|D|\delta_D$ is at most the norm of $|E|\delta_E$. First, note that

$$\delta_B - \kappa\delta_C = |B|^{-1}(|B|\delta_B - |C|\delta_C) = |B|^{-1}|B \setminus C|\delta_{B\setminus C} \ .$$

Now let D be a subset of $|B|$ of size between $\kappa_0|B_0|$ and $2\kappa_0|B_0|$ that contains $B \setminus C$. Let $|D| = \lambda|B|$. Then by our observations above, the norm of the right hand side is at most $|B|^{-1}|D|\|\delta_D\|_2 = \lambda\|\delta_D\|_2$. Because $|D| \geqslant \kappa_0|B_0|$, we know that $\|\delta_D\|_2^2 \leqslant (1+4\eta)\|d_B\|_2^2$. Therefore, we can certainly say that $\|\delta_B - \kappa\delta_C\|_2 \leqslant 2\lambda^{1/2}\|\delta_B\|_2$. We also know (using the fact that $\|\delta_C\|_2^2 \leqslant (1+4\eta)\|\delta_B\|_2^2$ and that $\eta \leqslant 1/4$), that

$$\|\delta_C - \kappa\delta_C\|_2 = (1-\kappa)\|\delta_C\|_2 \leqslant 2\kappa_0^{1/2}\|\delta_B\|_2 \ .$$

Since $\lambda \leqslant 4\kappa_0$, it follows from the triangle inequality that $\|\delta_C - \delta_B\|_2 \leqslant 6\kappa_0^{1/2}\|\delta_B\|_2$. Since $36\kappa_0^2 \leqslant \theta$, this shows what we wanted.

Now $\delta_B = \kappa\delta_C + (1-\kappa)\delta_{B\setminus C}$. It follows that $\delta_B - \delta_C = (1-\kappa)(\delta_{B\setminus C} - \delta_C)$, so our assumption implies that $\|\delta_{B\setminus C} - \delta_C\|_2 > \theta\|\delta_B\|_2$ as well. Let us use these facts to prove that $\kappa \leqslant 1/3$. Indeed,

$$\begin{aligned}
\|\delta_B\|_2^2 &= (1-\kappa)^2\|\delta_{B\setminus C}\|_2^2 + \kappa^2\|\delta_C\|_2^2 + 2\kappa(1-\kappa)\langle\delta_{B\setminus C},\delta_C\rangle \\
&= (1-\kappa)^2\|\delta_{B\setminus C}\|_2^2 + \kappa^2\|\delta_C\|_2^2 + \kappa(1-\kappa)\Big(\|\delta_{B\setminus C}\|_2^2 + \|\delta_C\|_2^2 - \|\delta_{B\setminus C} - \delta_C\|_2^2\Big) \\
&\leqslant \alpha(1+2\eta)\Big((1-\kappa)^2 + \kappa^2 + 2\kappa(1-\kappa)\Big) - \kappa(1-\kappa)\|\delta_{B\setminus C} - \delta_C\|_2^2 \\
&\leqslant \alpha(1+2\eta) - \kappa(1-\kappa)\theta\|\delta_B\|_2^2 \ .
\end{aligned}$$

If $\kappa > 1/3$, then $\kappa(1-\kappa) > 1/5$, so it follows that $\|\delta_B\|_2^2 < \alpha(1+2\eta)(1+\theta/5)^{-1}$. It can be checked that this is at most $(1-\eta)\alpha$, which contradicts the lower bound we have for $\|\delta_B\|_2^2$.

Those were just preliminaries, and we now come to the main part of the proof. From the second line of the above calculation we have

$$\|\delta_B\|_2^2 = (1-\kappa)\|\delta_{B\setminus C}\|_2^2 + \kappa\|\delta_C\|_2^2 - \kappa(1-\kappa)\|\delta_{B\setminus C} - \delta_C\|_2^2 \ .$$

By our assumption, this implies that

$$\big(1+\kappa(1-\kappa)\theta\big)\|\delta_B\|_2^2 \leqslant (1-\kappa)\|\delta_{B\setminus C}\|_2^2 + \kappa\|\delta_C\|_2^2 \ .$$

Now $\|\delta_C\|_2^2 \leqslant (1+2\eta)(1-\eta)^{-1}\|\delta_B\|_2^2 \leqslant (1+4\eta)\|\delta_B\|_2^2$, so

$$\big(1+\kappa(1-\kappa)\theta - (1+4\eta)\kappa\big)\|\delta_B\|_2^2 \leqslant (1-\kappa)\|\delta_{B\setminus C}\|_2^2 \ .$$

It can be checked that the left hand side is at least $(1-\kappa)(1+16\eta\kappa)\|\delta_B\|_2^2$, and also that $(1+16\eta\kappa) \geqslant (1-\kappa)^{-8\eta}$ when $\eta \leqslant 1/4$ and $\kappa < 1/3$, so this shows that $\|\delta_{B\setminus C}\|_2^2 \geqslant (1-\kappa)^{-8\eta}\|\delta_B\|_2^2$.

This allows us to do an iteration. We have just shown that if B is any subset of B_0 of size at least $|B_0|/2$ for which the conclusion does not hold, then we can find $\kappa \leqslant 1/3$ and a subset $B' \subset B$ of size $(1-\kappa)|B|$ such that $\|\delta_{B'}\|_2^2 \geqslant (1-\kappa)^{-8\eta}\|\delta_B\|_2^2$. Let us therefore choose a maximal sequence $B_0 \supset B_1 \supset \ldots \supset B_r$ such that $|B_i| = (1-\kappa_i)|B_{i-1}|$, $0 < \kappa_i \leqslant 1/3$ and $\|\delta_{B_i}\|_2^2 \geqslant (1-\kappa_i)^{-8\eta}\|\delta_{B_{i-1}}\|_2^2$ for each i.

If we do this, then there cannot be any k such that $\prod_{i=1}^k(1-\kappa_i) \leqslant 1/2$, since if we choose a minimal such k, then $1/2 \leqslant \prod_{i=1}^k(1-\kappa_i) \leqslant 3/4$. The upper bound and the regularity assumption imply that $\|\delta_{B_k}\|_2^2 \leqslant (1+4\eta)\|\delta_B\|_2^2$, while we also know that $\|\delta_{B_k}\|_2^2 \geqslant 2^{8\eta}\|\delta_B\|_2^2$. It can be checked that $8\log 2 > 4$, from which it follows (by looking at derivatives) that $2^{8\eta} > 1+4\eta$ for every $\eta > 0$.

This is a contradiction, so we can deduce that $\prod_{i=1}^k(1-\kappa_i) \leqslant 1/2$. This means that the iteration stops with a set B_r of size at least $|B_0|/2$ for which the conclusion does hold, which proves the lemma. $\square$

We need one more preliminary lemma, considerably easier than the previous one.

Lemma 5.3. *Let $0 < \alpha$, $0 < \eta \leqslant 1/4$, $0 < \gamma < \eta/4$ and $0 < \epsilon \leqslant \alpha^2\eta/4$. Let $\theta = 24\eta$ and $\kappa_0 = 2\alpha^{-1}(\epsilon/\eta)^{1/2}$. Let G be a bipartite graph with vertex sets X and Y, and suppose that*

$B \subset X$ is (ϵ, γ, η)-regular for G with parameter α. Let C be a subset of B of cardinality at least $\kappa_0|B|$ and suppose that $\|\delta_C - \delta_B\|_2^2 \leqslant \theta\|\delta_B\|_2^2$. Then $\mathbb{E}_{x,x' \in C}\Big|\delta_1(x,x') - \|\delta_C\|_2^2\Big|^2 \leqslant 5\theta^{1/2}\alpha^2$.

Proof. Since $|C| \geqslant \kappa_0|B|$, there are at most $\kappa_0^{-2}\epsilon|C|^2$ pairs $(x, x') \in C^2$ such that $\delta_1(x,x') \geqslant \alpha(1+\eta)$. Therefore, $\mathbb{E}_{x,x'\in C}\delta_1(x,x')^2 \leqslant \alpha^2(1+\eta)^2 + \kappa_0^{-2}\epsilon \leqslant \alpha^2(1+3\eta)$. We also know that $\mathbb{E}_{x,x'\in C}\delta_1(x,x') = \|\delta_C\|_2^2 \geqslant (1-\theta^{1/2})^2\|\delta_B\|_2^2$, by the triangle inequality and our assumption about C. We saw in Lemma 5.2 that $\|\delta_B\|_2^2 \geqslant \alpha(1-\eta)$, so

$$\mathbb{E}_{x,x'\in C}\Big|\delta_1(x,x') - \|\delta_C\|_2^2\Big|^2 \leqslant \alpha^2\big(1 + 3\eta - (1-\theta^{1/2})^4(1-\eta)^2\big)\ .$$

It can be checked that this is at most the bound stated. □

We now have the tools in place to prove the main theorem of this section.

Theorem 5.4. *Let N be a prime and let $A \subset \mathbb{Z}_N$ be a set of cardinality $\lfloor N/2 \rfloor$. Then there exists x such that $\Big||A \cap (A+x)| - N/4\Big| \leqslant AN(\log\log N)^{-c}$, where A and $c > 0$ are absolute constants.*

Proof. Define a bipartite graph G with vertex sets $X = Y = \mathbb{Z}_N$ by joining x to y if and only if $y - x \in A$. That is, join $x \in X$ to all points y of the form $x + a$. Notice first that if B is a subset of X and μ_B is the "characteristic measure" of B—that is, $\mu_B(x) = |B|^{-1}$ when $x \in B$ and 0 otherwise—then $\delta_B(y)$, which is the proportion of $x \in B$ such that $y - x \in A$, is $|B|^{-1}\sum_{x\in\mathbb{Z}_N} B(x)A(y-x)$, so $\delta_B = \mu_B * A$.

Let $0 < \eta \leqslant 1/4$, let $0 < \gamma \leqslant \eta/16$ and let $\epsilon = \eta/1024$. Now use Lemma 5.1 to choose B of density $\beta \geqslant \exp\exp(10\eta^{-1}\log\eta^{-1})$ that is $(\epsilon\beta^2, \gamma, \eta)$-regular with some parameter $\alpha \geqslant 1/4$. Then use Lemma 5.2 to pass to a subset B' of density $\beta' \geqslant \beta/2$, which is $(16\epsilon\beta'^2, 4\gamma, \eta)$-regular with parameter α (so far this is true for any subset B' of density $\beta' \geqslant \beta/2$) and which also has the property that $\|\delta_C - \delta_{B'}\|_2^2 \leqslant 24\eta\|\delta_{B'}\|_2^2$ for any subset $C \subset B'$ of size at least $(4\epsilon/\eta)^{1/2}\beta'|B'|$.

We now apply this to a special choice of C. The average size of $B' \cap (B' - d)$ over all $d \in \mathbb{Z}_N$ is $\beta'^2 N = \beta'|B|$. Therefore, we can find $d \neq 0$ such that $C = B' \cap (B' - d)$ has size at least $\beta'|B'|/2$, and so, therefore, does $C + d = B' \cap (B' + d)$. But by the triangle inequality and the property we have established for B', we know that $\|\delta_C - \delta_{C+d}\|_2^2 \leqslant 96\eta\|\delta_{B'}\|_2^2$. Since A has a fixed density we may as well forget the slight extra strength we have and use the fact that $\|\delta_{B'}\|_2^2 \leqslant 1$. Then, by our earlier remarks, $\|\mu_C * A - (\mu_C * A + d)\|_2^2 \leqslant 96\eta$. That is, loosely speaking, the function $\mu_C * A$ is approximately unchanged (in $L_2(X)$) if you translate it by d.

We now use a discrete intermediate value theorem to find x such that $|\langle\mu_C * A, \mu_C * A + x\rangle - 1/4| \leqslant 10\eta^{1/2}$. Observe first that the expectation of $\langle\mu_C * A, \mu_C * A + x\rangle$ is $\|\mu_C * A\|_1^2 = s = 1/4 + O(N^{-1})$. There must exist $x \neq x'$ such that $\langle\mu_C * A, \mu_C * A + x\rangle \leqslant s$ and $\langle\mu_C * A, \mu_C * A + x'\rangle \geqslant s$. Now write $x' = x + md$ for some m. (This is where we use the fact that N is prime.) Then there must be some minimal k such that $\langle\mu_C * A, \mu_C * A + x + kd\rangle \geqslant s$. So either $k = 0$ in which case we are clearly done, or $k \geqslant 1$ in which case $\langle\mu_C * A, \mu_C * A + x + (k-1)d\rangle < s$. But $\|(\mu_C * A + x + kd) - (\mu_C * A + x + (k-1)d)\|_2 \leqslant 10\eta^{1/2}$, by the approximate translation-invariance of $\mu_C * A$, and $\|\mu_C * A\| \leqslant 1$, so in the second case we must have $\langle\mu_C * A, \mu_C * A + x + kd\rangle \leqslant s + 10\eta^{1/2}$. Let us write $z = x + kd$.

To complete the proof we shall use the previous lemma and Lemma 4.4. The set B' satisfies the assumptions of that lemma if we replace ϵ by $16\epsilon\beta'^2$ and γ by 4γ. It therefore tells us that $\mathbb{E}_{x,x' \in C}\Big|\delta_1(x, x') - \|\delta_C\|_2^2\Big|^2 \leqslant 5\theta^{1/2}\alpha^2$ whenever C has size at least $4(16\epsilon\beta'^2/\eta)^{1/2}|B'| = \beta'|B'|/2$.

This gives us the hypotheses of Lemma 4.4 with $X = C$, $X' = C + z$, and $c = 10\theta^{1/2}$. (Here we are using the fact that $\|f\|_{U_2} = \|f'\|_{U_2} \geqslant \|\delta_C\|_2 \geqslant 1/2 + O(N^{-1}$.) It follows from that lemma that the variance of the quantity $\delta_1(x, x')$ over $(x, x') \in C \times (C + z)$ is at most $32\theta^{1/32}$. The mean is $\langle\delta_C, \delta_{C+z}\rangle$, which we have shown lies between s and $s + 10\eta^{1/2}$. Therefore, there exist $x \in C$ and $x' \in C + z$ such that $|\delta_1(x, x') - s| \leqslant 10\eta^{1/2} + 6\theta^{1/64}$.

Since we can choose η, and hence θ, to be as small as we like, this proves the result in a qualitative sense. To obtain the dependence on N, the main constraint is that B' should have a non-zero difference that occurs at least $\beta'|B'|/2$ times. For this it is sufficient if the density β of B is at least $N^{-1/3}$, say. This forces η to be at least $c \log\log\log N/\log\log N$, and therefore forces $10\eta^{1/2} + 6\theta^{1/64}$ to be at least $c(\log\log\log N/\log\log N)^{1/64}$. □

§6. Concluding remarks.

In one respect the weak regularity lemma that we obtained in Theorem 4.3 is a little artificial. To see why, let X_1, X_2, Y_1 and Y_2 be disjoint finite sets, all of the same size, and form a bipartite graph G by joining $x \in X_i$ to $y \in Y_j$ randomly with probability p_{ij}, where $p_{ij} = 1/2$ if $i = j$ and $1/4$ if $i \neq j$.

It is not hard to check that if x and x' belong to the same X_i, then with high probability $\delta_1(x, x')$ is approximately $5/32$, whereas if they belong to different X_i then with high probability $\delta_1(x, x')$ is approximately $1/8$. In fact, the quantifiers can be reversed: with high probability the graph is such that these are good approximations for *all* the pairs. It is clear, therefore, that the "natural" partition of X is into the sets X_1 and X_2: this will have all the properties we could possibly want from any regularizing partition of X.

However, if we use the method of proof of Theorem 4.3, we will find ourselves applying Corollary 4.2, which itself uses Lemma 4.1, and when we examine how Lemma 4.1 identifies good subsets of X, we see that it produces sets that are, quite unnecessarily, much smaller than X_1 and X_2. The basic idea behind Lemma 4.1 is to choose random points $y_1, \ldots, y_k$ in Y and to take the intersection of their neighbourhoods. If we do that in this example, then the proof works for the following rough reason: although the probability of choosing any given x is rather small, the probability that we choose x' given that we choose x is far larger if x and x' belong to the same X_i than if they belong to different X_i.

This correlation effect becomes stronger and stronger as k increases, but so does the probability of choosing any given x, and therefore so does the expected size of the set we eventually pick. In fact, for this graph even the maximum size of the common neighbourhood of k vertices in Y decreases exponentially in k.

We therefore end up choosing very small subsets of X_1 and X_2, and the partition in Theorem 4.3 is (approximately, at any rate) a refinement of the partition $\{X_1, X_2\}$ into a large number of much smaller sets, all of which behave in an almost identical way to the X_i of which they are (more or less) a subset.

Theorem 4.7, our "GG^* regularity lemma," addresses this problem to some extent, because it gives us a way of conglomerating the cells of the partition when it is appropriate to do so: if K is a set such that the vectors δ_{B_i} with $i \in K$ are all close to each other, then all the inner products $\langle \delta_{B_i}, \delta_{B_j} \rangle$ with $i, j \in K$ are close to each other. Moreover, they are all close to $\|\delta_B\|_2^2$, where $B = \bigcup_{i \in K} B_i$. Therefore, this union has a regularity property as well.

However, as the above argument demonstrates, in some ways it is more natural to look for a *metric* on G than a partition, where we use the word "metric" slightly loosely in that we allow $d(x, y)$ to equal 0 even if $x \neq y$. (In other words, we should call it a pseudometric, but since what will really interest us is the triangle inequality we have not bothered to do so.) To get a further idea of why this might be, consider the "sphere graph" that has the n-sphere S_n as its vertex set, and joins x to y if $\langle x, y \rangle \geqslant 0$. Then the size of the intersection of the neighbourhoods of x and y is a decreasing function of their distance, so this distance is quite clearly more natural than any partition of the vertex set, since the latter will necessarily create artificial boundaries.

Another nice way of putting a metric on X is to define $d(x, x')$ to be the distance in

$L_2(Y)$ between the unit vectors $\delta_{B_i}/\|\delta_{B_i}\|_2$ and $\delta_{B_j}/\|\delta_{B_j}\|_2$, where $x \in B_i$ and $x' \in B_j$. This time, if x and x' are close, all we can say is that δ_{B_i} and δ_{B_j} are roughly proportional. However, the constant of proportionality is equal to the ratio of the densities of the graphs $G(B_i, Y)$ and $G(B_j, Y)$, and from that one can show, by means of a proof very similar to that of Theorem 4.7, that if we take a small neighbourhood B in this metric, then the graph $G(B, Y)$ will be approximately of rank 1 in the sense of Corollary 3.8 and Theorem 3.11. We omit the details.

References.

[A] N. Alon, *Eigenvalues and expanders*, Combinatorica **6** (1986), 83-96.

[AM] N. Alon, V. D. Milman, *Isoperimetric inequalities for graphs and superconcentrators*, J. Combin. Theory Ser. B **38** (1985), 73-88.

[BS] A. Balog, E. Szemerédi, *A statistical theorem of set addition*, Combinatorical **14** (1994), 263-268.

[CG] F. R. K. Chung, R. L. Graham, *Quasi-random graphs with given degree sequences*, Random Structures and Algorithms, to appear.

[CGW] F. R. K. Chung, R. L. Graham, R. M. Wilson, *Quasi-random graphs*, Combinatorica **9** (1989), 345–362.

[G1] W. T. Gowers *Lower bounds of tower type for Szemerédi's uniformity lemma*, Geom. Funct. Anal. **7** (1997), 322-337.

[G2] W. T. Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. **11** (2001), 465-588.

[G3] W. T. Gowers, *Quasirandomness, counting and regularity for 3-uniform hypergraphs*, Combin. Probab. Comput. **15** (2006), 143-184.

[G4] W. T. Gowers, *Quasirandom groups* (2006), submitted.

[GK] B. J. Green, S. Konyagin, *On the Littlewood problem modulo a prime*, Canadian J. Math., to appear.

[S] T. Sanders, *The Littlewood-Gowers problem*, J. Anal. Math., to appear.

[T] A.G. Thomason, *Pseudo-random graphs*, Proceedings of Random Graphs, Poznán 1985 (M. Karonski, ed.), Annals of Discrete Mathematics **33**, 307–331.