The Effect of Induced Subgraphs on Quasi-Randomness

Asaf Shapira * Raphael Yuster †

Abstract

One of the main questions that arise when studying random and quasi-random structures is which properties $\mathcal{P}$ are such that any object that satisfies $\mathcal{P}$ "behaves" like a truly random one. In the context of graphs, Chung, Graham, and Wilson [9] call a graph *p-quasi-random* if it satisfies a long list of the properties that hold in $G(n,p)$ with high probability, like edge distribution, spectral gap, cut size, and more.

Our main result here is that the following holds for **any** fixed graph H: if the distribution of *induced* copies of H in a graph G is close (in a well defined way) to the distribution we would expect to have in $G(n,p)$, then G is either p-quasi-random or $\overline{p}$-quasi-random, where $\overline{p}$ is the unique non-trivial solution of the polynomial equation $x^{\delta}(1-x)^{1-\delta} = p^{\delta}(1-p)^{1-\delta}$, with δ being the edge density of H. We thus infer that having the correct distribution of induced copies of any single graph H, is enough to guarantee that a graph has the properties of a random one. The proof techniques we develop here, which combine probabilistic, algebraic and combinatorial tools, may be of independent interest to the study of quasi-random structures.

1 Introduction

1.1 Background and basic definitions

Quasi-random (or pseudo-random) structures are those that possess the properties we expect random objects to have with high probability. The study of quasi-random structures is one of the most interesting borderlines between discrete mathematics and theoretical computer science, as they relate the problem of how to deterministically construct a random-like object with the question of when can we consider a single event to be a random one. Although quasi-random structures have been implicitly studied for many decades, they were first explicitly studied in the context of graphs by Thomason [27, 28] and then followed by Chung, Graham, and Wilson [9]. Following the results on quasi-random graphs, quasi-random properties were also studied in various other contexts such as set systems [5], tournaments [6], and hypergraphs [7]. There are also some very recent results on quasi-random groups [11] and generalized quasi-random graphs [16]. We briefly mention that the study of quasi-random structures lies at the core of the recent proofs of Szemerédi's Theorem [24] that were recently obtained independently by Gowers [12, 13] and by Rödl et. al. [19, 17] and then also by Tao [26]. For more mathematical background on quasi-randomness the reader is referred to the recent papers of Gowers [11, 12, 13] and to the survey of Krivelevich and Sudakov [15].

*Microsoft Research, Redmond. E–mail: asafico@tau.ac.il

†Department of Mathematics, University of Haifa, Haifa 31905, Israel. E–mail: raphy@research.haifa.ac.il

Besides being intriguing questions on their own, results on quasi-random objects also have applications in theoretical computer science. The main point is that while the classical definitions of what it means for an object to be quasi-random are hard to verify, some other properties, which can be proved to be equivalent, are much easier to verify. The archetypal example of this phenomena is probably the spectral gap property of *expanders*. Expanders are sparse graphs that behave like random sparse graphs in many aspects (see [14] for more details), and are one of the most widely used structures in theoretical computer science. However, verifying that a graph satisfies the classical definition of being an expander, that is, that any cut has many edges, requires exponential time. A very useful fact is that being an expander is equivalent to the fact that the second eigenvalue of the adjacency matrix of the graph is significantly smaller than the first eigenvalue (see also Property $\mathcal{P}_3$ in Theorem 1). As eigenvalues can be computed in polynomial time, this gives an efficient way to verify that a sparse graph is an expander. Another example, this time on dense graphs, is that a natural notion of quasi-randomness for a dense graphs is that all subsets of vertices should contain the "correct" number of edges as in $G(n,p)$. This property takes exponential time to verify, but fortunately (see Theorem 1), it turns out that this property is equivalent to the property of having the "correct" number of edges and copies of the cycle of length four in the entire graph! As this property takes only polynomial time to verify, this gives an efficient algorithm for checking if a dense graph is quasi-random. This easily verifiable condition was a key (implicit) ingredient in the work of Alon et al. [1] who gave the first polynomial time algorithm for Szemerédi's Regularity Lemma [25], whose original proof was non-constructive.

Given the above discussion, one of the most natural questions that arise when studying quasi-random objects, is which properties "guarantee" that an object behaves like a truly-random one. Our main result in this paper establishes that for any single graph H, if the distribution of the induced copies of H in a graph G is "close", in some precise sense, to the one we expect to have in $G(n,p)$, then G is quasi-random. Previous studies [22, 23] of the effect of induced subgraph on quasi-randomness that used a slightly weaker notion of "closeness", indicated that in some cases the distribution of induced copies of a single graph H is *not enough* to guarantee that a graph is quasi-random. Therefore, the notion of closeness that we use here is essentially optimal if one wants to be able to deal with any H.

Before stating our main result we first discuss some previous ones, which will put ours in the right context. The cornerstone result on properties guaranteeing that a graph is quasi-random is that of Chung, Graham, and Wilson [9], stated below, but before stating it we need to introduce some notation. We will denote by $e(G)$ the number of edges of a graph G. A *labeled copy* of a graph H in a graph G is an injective mapping ϕ, from the vertices of H to the vertices of G, that maps edges to edges, that is $(i,j) \in E(H) \Rightarrow (\phi(i),\phi(j)) \in E(G)$. So the expected number of labeled copies of a graph H in $G(n,p)$, is $p^{e(H)}n^h + o(n^h)$ where h is the number of vertices of H [1]. A *labeled induced copy* of a graph H in a graph G is an injective mapping ϕ, from the vertices of H to the vertices of G, that maps edges to edges, and non edges to non edges, that is

[1]Note that this is *not* the expected number of (unlabeled) copies of H in G, which is just the number of labeled copies of H divided by the number of automorphisms of H. Therefore, all the result we mention here also hold when considering (unlabeled) copies. We work with labeled copies (induced or not) because we do not need to refer to the automorphisms of H, and because it is easier to count labeled copies than copies.

$(i,j) \in E(H) \Leftrightarrow (\phi(i), \phi(j)) \in E(G)$. So the expected number of induced labeled copies of a graph H in $G(n,p)$, is $\delta_H(p)n^h + o(n^h)$, where here and throughout the paper we will use $\delta_H(p)$ to denote $p^{e(H)}(1-p)^{\binom{h}{2}-e(H)}$. For a set of vertices $U \subseteq V$ we denote by $H[U]$ the number of labeled copies of H in U, and by $H^*[U]$ the number of induced labeled copies of H in U. The following is (part of) the main result of [9]:

Theorem 1 (Chung, Graham, and Wilson [9]) *Fix any $1 < p < 1$. For any n-vertex graph G the following properties are equivalent:*

$\mathcal{P}_1$: *For any subset of vertices $U \subseteq V(G)$ we have $e(U) = \frac{1}{2}p|U|^2 + o(n^2)$.*

$\mathcal{P}_2$: *For any subset of vertices $U \subseteq V(G)$ of size $\frac{1}{2}n$ we have $e(U) = \frac{1}{2}p|U|^2 + o(n^2)$.*

$\mathcal{P}_3$: *Let $\lambda_i(G)$ denote the i^{th} largest (in absolute value) eigenvalue of G. Then $e(G) = \frac{1}{2}pn^2 + o(n^2)$, $\lambda_1(G) = pn + o(n)$ and $\lambda_2(G) = o(n)$.*

$\mathcal{P}_4(t)$: *For an even integer $t \geq 4$, let C_t denote the cycle of length t. Then $e(G) = \frac{1}{2}pn^2 + o(n^2)$ and $C_t[G] = p^t n^t + o(n^t)$.*

$\mathcal{P}_5$: *Fix an $\alpha \in (0, \frac{1}{2})$. For any $U \subseteq V(G)$ of size αn we have $e(U, V \setminus U) = p\alpha(1-\alpha)n^2 + o(n^2)$.*

The meaning of the fact that, for example, $\mathcal{P}_2$ *implies* $\mathcal{P}_1$ is that for any $\delta > 0$ there is an $\epsilon = \epsilon(\delta)$ such that if G has the property that all $U \subseteq V(G)$ of size $n/2$ satisfy $e(U) = \frac{1}{2}p|U|^2 \pm \epsilon n^2$, then $e(U) = \frac{1}{2}p|U|^2 \pm \delta n^2$ for all $U \subseteq V(G)$ [2]. This will also be the meaning of other *implications* between other graph properties later on in the paper. Here and throughout the paper, $x = y \pm \epsilon$ is shorthand for $y - \epsilon \leq x \leq y + \epsilon$.

Note, that each of the items in Theorem 1 is a property we would expect $G(n,p)$ to satisfy with high probability. We will thus say that G is *p-quasi-random* if it satisfies property $\mathcal{P}_1$, that is if for some small δ all $U \subseteq V(G)$ satisfy $e(U) = \frac{1}{2}p|U|^2 \pm \delta n^2$. If one wishes to be more precise then we can in fact say that such a graph is (p,δ)-quasi-random. We will sometimes omit the p and just say that a graph is quasi-random. In the rest of the paper the meaning of a statement "If G satisfies $\mathcal{P}_2$ then G is quasi-random" is that $\mathcal{P}_2$ implies $\mathcal{P}_1$ in the sense of Theorem 1 discussed in the previous paragraph. We will also say that a graph property $\mathcal{P}$ is *quasi-random* if any graph that satisfies $\mathcal{P}$ must be quasi-random. So the meaning of the statement "$\mathcal{P}_2$ is quasi-random" is that $\mathcal{P}_2$ implies $\mathcal{P}_1$. Therefore, all the properties in Theorem 1 are quasi-random.

Given Theorem 1 one may stipulate that any property that holds with high probability in $G(n,p)$ is quasi-random. That however, is far from true. For example, it is easy to see that having the "correct" vertex degrees is not a quasi-random property (consider $K_{n/2,n/2}$). Note also that in $\mathcal{P}_5$ we require $\alpha < \frac{1}{2}$, because when $\alpha = \frac{1}{2}$ the property is not quasi-random (see [8] and [22]). A more relevant family of non quasi-random properties are those requiring the graph to have the correct number of copies of a fixed graph H. Note that $\mathcal{P}_4(t)$ guarantees that for any even t, if a graph has the correct number of edges and the correct number of copies of C_t then it is quasi-random. As observed in [9] this is *not* true for all graphs, in fact this is not true for any non-bipartite H.

[2] An equivalent, more cumbersome, way to state Theorem 1 would have been to replace all the $o(.)$ terms by $\epsilon_1, \ldots, \epsilon_5$ and say that there are some functions $f_{i,j}$ that relate ϵ_i and ϵ_j.

1.2 Quasi-randomness and the distribution of copies of a single graph

As throughout the paper we work with labeled copies and labeled induced copies of H, we henceforth just call them copies and induced copies. To understand the context of our main result that deals with induced copies of a fixed graph H, it is instructive to review what is known about the effect of the distribution of a fixed graph H on quasi-randomness. By Theorem 1 we know that for some graphs H the property of having the correct number of copies of H in G, along with the right number of edges, is enough to guarantee that G is quasi-random. Furthermore, this is not true for all graphs H. However, the intuition is that something along these lines should be true for any H, that is that for any H, if the copies of H in a graph G have the "properties" we would expect them to have in $G(n, p)$, then G should be p-quasi-random. Simonovits and Sós [22] observed that the counter examples showing that for some graphs H, having just the correct number of copies of H (and the correct number of edges) is not enough to guarantee quasi-randomness, all have the property that some of the induced subgraphs of these counter examples have significantly more/less copies of H than we would expect to find in $G(n, p)$. For example, in order to show that having the correct number of edges and triangles as in $G(n, 1/2)$ does not guarantee that G is $\frac{1}{2}$-quasi-random, one can take a complete graph on αn vertices and a complete bipartite graph on $(1 - \alpha)n$ vertices, for an appropriate α.

The main insight of Simonovits and Sós [22] was that quasi-randomness is a *hereditary* property, in the sense that we expect a sub-structure of a random-like object to be random-like as well. Thus, perhaps it will suffice to require that the subgraphs of G should also have the correct number of copies of H. To state the main result of [22] let us introduce the following variant of property $\mathcal{P}_1$ of Theorem 1.

Definition 1.1 **($H[U_1, \ldots, U_h]$)** *For a graph H on h vertices, and pairwise disjoint vertex sets $U_1, \ldots, U_h$, we denote by $H[U_1, \ldots, U_h]$ the number of h-tuples $v_1 \in U_1, \ldots, v_h \in U_h$ that span a labeled copy of H.*

Definition 1.2 **($\mathcal{P}_H$)** *For a fixed graph H on h vertices, we say that a graph G satisfies $\mathcal{P}_H$ if all pairwise disjoint h-tuples $U_1, \ldots, U_h \subseteq V(G)$ of equal (arbitrary) size m satisfy*

$$H[U_1, \ldots, U_h] = p^{e(H)} h! m^h + o(n^h) \ .$$

Note that the above restriction is that the value of $H[U_1, \ldots, U_h]$ should be close to what it should be in $G(n, p)$ for all h-tuples of equal-size. Observe also that the above condition does not impose any restriction on the number of edges of G, while in property $\mathcal{P}_1$ there is. Note also, that the error in the above definition involves n rather than $m = |U_1| = \cdots = |U_h|$ so when $m = o(n)$ the condition vacuously holds. As opposed to $\mathcal{P}_4$, which is not quasi-random for all graphs, Simonovits and Sós [22] showed that $\mathcal{P}_H$ is quasi-random for any graph H.

Theorem 2 (Simonovits and Sós [22]) *The following holds for any graph H: if a graph G satisfies $\mathcal{P}_H$ then it is p-quasi-random.*

Observe that $\mathcal{P}_H$ requires, via Definition1.1, all h-tuples of vertex sets to have the correct number of copies of H with one vertex in each set. A more "natural" requirement, that was actually used in [22], is that all subsets of vertices $U \subseteq V(G)$ should contain the correct number of copies of H, that is, that $H[U] \approx p^{e(H)}|U|^h$ for all $U \subseteq V(G)$. However, it is not difficult to show that these two conditions are in fact equivalent (see [20]). We choose to work with Definition 1.1 as it will fit better with the discussion in the next subsection.

1.3 The main result

So we know from Theorem 1 that when we consider the number of subgraphs of H in G, then some H but not all, are such that having the correct number of copies of H in a graph G (and number of edges) is enough to guarantee that G is quasi-random. From Theorem 2 we know that for all H, having the correct number of copies of H in all the subgraphs of G is enough to guarantee that G is quasi-random. A natural question is what can we learn from the distribution of *induced* copies of a graph H? As we shall see, the situation is much more involved.

Recall that for a fixed graph H on h vertices and a fixed $0 < p < 1$, we define $\delta_H(p) = p^{e(H)}(1-p)^{\binom{h}{2}-e(H)}$. Let us denote by $\overline{p}_H$ the second [3] solution (other than p) of the equation $\delta_H(p) = x^{e(H)}(1-x)^{\binom{h}{2}-e(H)}$. We call $\overline{p}_H$ the *conjugate* of p with respect to H. We will sometimes just write $\overline{p}$ instead of $\overline{p}_H$ when H is fixed. Note that the expected number of induced copies of H in a set of vertices U is roughly $\delta_H(p)|U|^h$. But, as it may [4] be the case that $p \neq \overline{p}_H$ we see that for any H and any p, the distribution of induced copies of H in both $G(n,p)$ and $G(n,\overline{p}_H)$ behaves *precisely* the same. Therefore, the best we can hope to deduce from the fact that the distribution of induced copies of H in G is close to that of $G(n,p)$ is that G is either p-quasi-random or $\overline{p}_H$-quasi-random.

Let us denote by $H^*[U_1,\ldots,U_h]$ the natural generalization of $H[U_1,\ldots,U_h]$ (defined in Definition 1.1) with respect to induced subgraphs, that is, $H^*[U_1,\ldots,U_h]$ is the number of h-tuples of vertices $v_1 \in U_1,\ldots,v_h \in U_h$ with the property that $v_1,\ldots,v_h$ span a labeled induced copy of H. Note that for an h tuple of vertex sets $U_1,\ldots,U_h$ in $G(n,p)$ each of size m, the expected value of $H^*[U_1,\ldots,U_h]$ is $\delta_H(p)h!m^h$.

So given the above discussion and Theorem 2, it seems reasonable to conjecture that if a graph G has the correct distribution of induced copies of H, then G is either p-quasi-random or $\overline{p}_H$-quasi-random. When we say correct distribution we mean that all pairwise disjoint h-tuples $U_1,\ldots,U_h \subseteq V(G)$ of the same size m satisfy $H^*[U_1,\ldots,U_h] \approx \delta_H(p)h!m^h$. However, it was observed in [22, 23] that this is not the case. For example, one can take vertex set V_1, V_2 of sizes $\alpha n, (1-\alpha)n$ and put $G(\alpha n, p_1)$ on V_1, $G((1-\alpha_1)n, p_1)$ on V_2 and connect V_1 and V_2 with probability $p_2 \neq p_1$. Then for appropriate constants, we get a graph with the correct distribution of the 3-vertex path, yet this graph is not p-quasi-random for any p.

However, as before, the intuition is that having the correct distribution of induced copies of H *should* guarantee that G is quasi-random. Our main result in this paper is that indeed it does, one just needs to refine the notion of "correct distribution". As we have mentioned before, if $U_1,\ldots,U_h$ is an h-tuple of vertices in $G(n,p)$ of the same size m, then we would expect to have $H^*[U_1,\ldots,U_h] \approx \delta_H(p)h!m^h$. But observe, that the reason for that, is that we would actually

[3] It is not difficult to see that for non-negative integers k, ℓ the equation $x^k(1-x)^\ell = q$ has at most two solutions in $(0,1)$.

[4] The only case where $p = \overline{p}_H$ is when $p = e(H)/\binom{h}{2}$

expect a slightly stronger condition to hold. Before stating this condition, let us introduce the following "permuted" version of the quantity $H^*[U_1, \dots, U_h]$.

Definition 1.3 **($H^*_\sigma[U_1, \dots, U_h]$)** *Let H be a graph on h vertices, let $U_1, \dots, U_h$ be an h-tuple of pairwise disjoint vertex sets, and let $\sigma \in S_h$ be permutation $[h] \to [h]$. Then we denote by $H^*_\sigma[U_1, \dots, U_h]$ the number of h-tuples of vertices $v_1 \in U_{\sigma(1)}, \dots, v_h \in U_{\sigma(h)}$ with the property that $v_i \in U_{\sigma(i)}$ is connected to $v_j \in U_{\sigma(j)}$ if and only if $(i,j) \in H$.*

Getting back to our discussion, observe that the reason we expect to have $H^*[U_1, \dots, U_h] \approx \delta_H(p)h!m^h$ is simply because we expect to have $H^*_\sigma[U_1, \dots, U_h] \approx \delta_H(p)m^h$ for all $h!$ permutations in S_h! It is now natural to define the following property:

Definition 1.4 **($\mathcal{P}^*_H$)** *For a fixed graph H on h vertices, we say that a graph G satisfies $\mathcal{P}^*_H$ if for all pairwise disjoint h-tuples $U_1, \dots, U_h \subseteq V(G)$ of equal (arbitrary) size m, and for every $\sigma \in S_h$*

$$H^*_\sigma[U_1, \dots, U_h] = \delta_H(p)m^h + o(n^h) \ .$$

Our main result is that property $\mathcal{P}^*_H$ guarantees that a graph is quasi-random.

Theorem 3 (Main result) *The following holds for any graph H: if a graph satisfies $\mathcal{P}^*_H$ then it is either p-quasi-random or $\overline{p}_H$-quasi-random.*

Our main result can be formulated as saying that for any H, if a graph G has the correct distribution of induced copies of H, then G is quasi-random. We remind the reader that one cannot hope to strengthen Theorem 3 by showing that G must be p-quasi-random, as $G(n, \overline{p})$ satisfies $\mathcal{P}^*_H$ with probability 1. Observe, that our notion of "correct distribution" (that is, the quantities H^*_σ) is just slightly stronger than the notions that have been considered before (that is, the quantities H^*), where the latter is known to be too weak to guarantee quasi-randomness.

1.4 Overview of the paper

As we have discussed in the first subsection, the theory of quasi-random graphs has many applications in theoretical computer science, both in the case of sparse and dense graphs. We think that the main interest of our result is in the proof techniques and tools that are used in the course of its proof. Besides several combinatorial arguments and tools (such as the Regularity Lemma [25], Ramsey's Theorem and Rödl's "nibble" Theorem [18]) the main underlying idea of the proof is an algebraic one. Roughly speaking, what we do is take all the information we know about the graph G, namely the information on the distribution of induced copies of H, and use it in order define a large system of polynomial equations. The unknowns in this system of equations represent (in some way) the distribution of edges of G. The crux of the proof is to show that the unique solution of this system of equations, is one that forces the edges of the graph to be nicely distributed (in the sense of property $\mathcal{P}_1$ in Theorem 1). The main theorem we need in order to obtain this uniqueness is a result of Gottlieb [3], in algebraic combinatorics, concerning the rank of set inclusion matrices (see Theorem 4). This approach to showing that a graph is quasi-random may be applicable for showing quasi-random properties of other structures.

In Section 2 we prove Theorem 3 by applying several combinatorial tools as well as a key lemma (Lemma 2.1) that is proved in Section 3. The proof of Lemma 2.1, which is the most difficult step in the proof of Theorem 3, contains most of the new ideas we introduce in this paper.

2 Proof of Main Result

In this section we give the proof of Theorem 3, but before getting to the actual proof we will need some preparation. We first discuss Lemma 2.1 which is the main technical lemma we need for the proof of Theorem 3, and whose proof appears in the next section. We then discuss some simple notions related to the Regularity Lemma, and then turn to the proof of Theorem 3. Throughout this section, let us fix a real $0 < p < 1$ and a graph H on h vertices. Recall that we set $\delta_H(p) = p^{e(H)}(1-p)^{\binom{h}{2}-e(H)}$ and that we denote by $\overline{p}$, the conjugate of p, the second solution in $(0,1)$ of the equation $\delta_H(p) = x^{e(H)}(1-x)^{\binom{h}{2}-e(H)}$.

2.1 The Key Lemma

In what follows we will work with weighted complete graphs W on r vertices. We will think of the vertices of W as the integers $[r]$. In that case each pair of vertices $1 \le i < j \le r$ will have a weight $0 \le w(i,j) \le 1$. Let us identify the h vertices of H with the integers $[h]$. Given an injective mapping $\phi : [h] \to [r]$, which we think of as a mapping from the vertices of H to the vertices of W, we will set

$$W(\phi) = \prod_{(i,j)\in E(H)} w(\phi(i),\phi(j)) \prod_{(i,j)\notin E(H)} (1 - w(\phi(i),\phi(j))) \ .$$

Another notation that will simplify the presentation is a variant of the $H^*_\sigma[U_1,\ldots,U_h]$ notation that was defined in Section 1. Suppose we have r pairwise disjoint vertex sets $U_1,\ldots,U_r$ and an injective mapping $\phi : [h] \to [r]$. Then we denote by $H^*_\phi[U_1,\ldots,U_r]$ the number of h-tuples of vertices $v_1 \in U_{\phi(1)},\ldots,v_h \in U_{\phi(h)}$ with the property that $v_i \in U_{\phi(i)}$ is connected to $v_j \in U_{\phi(j)}$ if an only if (i,j) is an edge of H.

Suppose we construct an r-partite graph on vertex sets $U_1,\ldots,U_r$, each of size m, by connecting every vertex in U_i with any vertex in U_j independently with probability $w(i,j)$. Then, observe that for any $\phi : [h] \to [r]$, we would expect $H^*_\phi[U_1,\ldots,U_r]$ to be close to $W(\phi)m^h$. Continuing this example, suppose that all (i,j) satisfy $w(i,j) = p$. Then we would expect all ϕ to satisfy $H^*_\phi[U_1,\ldots,U_h] = \delta_H(p)m^h$. Observe however, that we would also expect the same to hold if we were to replace p by $\overline{p}$.

The following lemma shows that the converse is also true in the following sense: if we know that for *any* injective mapping ϕ we have the correct fraction of induced copies of H as we would expect to find if we had $w(i,j) = p$ for all (i,j), then either[5] almost all (i,j) satisfy $w(i,j) = p$ or almost all satisfy $w(i,j) = \overline{p}$. Note that for convenience the lemma is stated with respect to quantities in

[5] Remember that we cannot expect to be able to show that all densities are p as the number of induced copies of H behaves the same with respect to p and $\overline{p}$.

$(0,1)$, rather than with respect to the number of edges or number of copies [6]. In what follows, we will always assume wlog that if $p \neq \overline{p}$ then $\epsilon < |p - \overline{p}|/2$. This will guarantee that $\overline{p} \pm \epsilon \neq p \pm \epsilon$.

Lemma 2.1 (The Key Lemma) *For every h there exists a $N_{2.1} = N_{2.1}(h)$ so that for any $r \geq N_{2.1}$ and $\epsilon > 0$ there exists $\delta_{2.1} = \delta_{2.1}(\epsilon, h, r) > 0$ with the following properties: suppose W is a weighted graph on r vertices, such that for all $\phi : [h] \to [r]$ we have $W(\phi) = \delta_H(p) \pm \delta_{2.1}$. Then any pair (i,j) satisfies either $w(i,j) = p \pm \epsilon$ or $w(i,j) = \overline{p} \pm \epsilon$. Furthermore, either at most $r-1$ of the pairs (i,j) satisfy $w(i,j) = p \pm \epsilon$ or at most $r-1$ of the pairs (i,j) satisfy $w(i,j) = \overline{p} \pm \epsilon$.*

The proof of Lemma 2.1, which is the main lemma we need for the proof of Theorem 3, appears in Section 3. It is interesting to note that as we show in Section 3, one cannot strengthen the above lemma by showing that either all densities are close to p or they are all close to $\overline{p}$.

2.2 The Regularity Lemma

We now give a brief overview of the Regularity Lemma of Szemerédi, which turns out to be strongly related to quasi-random graphs. For a pair of nonempty vertex sets (A, B) we denote by $d(A, B)$ the edge density between A and B, that is $d(A, B) = |E(A, B)|/|A||B|$. A pair of vertex sets (A, B) is said to be γ-*regular*, if for any two subsets $A' \subseteq A$ and $B' \subseteq B$, satisfying $|A'| \geq \gamma|A|$ and $|B'| \geq \gamma|B|$, the inequality $|d(A', B') - d(A, B)| \leq \gamma$ holds. A partition of the vertex set of a graph is called an *equipartition* if all the sets of the partition are of the same size (up to 1). We call the number of partition classes of an equipartition the *order* of the equipartition. Finally, an equipartition $\mathcal{V} = \{V_1, \ldots, V_k\}$ of the vertex set of a graph is called γ-*regular* if all but at most $\gamma\binom{k}{2}$ of the pairs (V_i, V_j) are γ-regular. The celebrated Regularity Lemma of Szemerédi can be formulated as follows:

Lemma 2.2 ([25]) *For every t and $\gamma > 0$ there exists $T = T_{2.2}(\gamma, t)$, such that any graph of size at least t has a γ-regular equipartition of order k, where $t \leq k \leq T$.*

The following lemma of Simonovits and Sós [21] shows that the property of having a regular partition where most of the pairs are connected by regular pairs of density close to p implies that the graph is p-quasi-random. For completeness we include a short self contained proof of this lemma at the end of this section.

Lemma 2.3 (Simonovits and Sós [21]) *For every $\zeta > 0$ there is an $\epsilon = \epsilon_{2.3}(\zeta)$ and $t = t_{2.3}(\zeta)$ with the following property: suppose an n vertex graph G has an ϵ-regular partition of order $k \geq t$ where all but $\epsilon\binom{k}{2}$ of the pairs are ϵ-regular with density $p \pm \epsilon$. Then every set of vertices $U \subseteq G$ spans $\frac{1}{2}p|U|^2 \pm \zeta n^2$ edges.*

Another tool we will need for the proof of Theorem 3 is Lemma 2.4 below. This lemma is equivalent to saying that if we have r pairwise disjoint vertex sets $V_1, \ldots, V_r$ that are all regular enough, then for any injective mapping $\phi : [h] \to [r]$ we have that $H^*_\phi[V_1, \ldots, V_r]$ is close to what it should be. Such a lemma is well known, and has been proven and used in many papers. See, e.g., Lemma 4.2 in [10] for one such proof. We thus omit the proof of Lemma 2.4.

[6] So $w(i,j)$ should be understood as the density between the pair (U_i, U_j) and $W(\phi)$ is $H^*_\phi[U_1, \ldots, U_r]/m^h$.

Lemma 2.4 *For any $\delta > 0$ and h, there exists a $\gamma = \gamma_{2.4}(\delta, h) > 0$ such that the following holds: Let W be a weighted complete graph on r vertices, and suppose $V_1, \ldots, V_r$ are pairwise disjoint sets of vertices of size m each, that all pairs (V_i, V_j) are γ-regular and that all pairs satisfy $d(V_i, V_j) = w(i, j)$. Then, for any injective mapping $\phi : [h] \to [r]$, we have*

$$H^*_\phi[V_1, \ldots, V_r] = (W(\phi) \pm \delta)m^h. \tag{1}$$

2.3 Proof of Theorem 3

For the proof of Theorem 3 we will also need the following two lemmas, whose proof is deferred to the end of this section.

Lemma 2.5 *For every ϵ there is an $r_{2.5} = r_{2.5}(\epsilon)$ such that for every $r \geq r_{2.5}$ there is $N_{2.5} = N_{2.5}(r)$ and $\gamma_{2.5} = \gamma_{2.5}(r)$ with the following property. Assume $k \geq N_{2.5}$ and that K is a k vertex graph with at least $(1 - \gamma_{2.5})\binom{k}{2}$ edges. Suppose the edges of K are colored red/blue such that at least $\epsilon\binom{k}{2}$ are blue and at least $\epsilon\binom{k}{2}$ are red. Then K has r vertices that span a complete graph K_r with at least r blue edges and at least r red edges.*

Lemma 2.6 *For any γ and r, there is an $N_{2.6} = N_{2.6}(\gamma, r)$ such that the following holds for any $k \geq N_{2.6}$. If K is a graph on k vertices with at least $(1-\gamma)\binom{k}{2}$ edges, then K has at least $(1-\gamma r^2)\binom{k}{2}$ edges that belong to a copy of K_r.*

Proof of Theorem 3: We will say that a γ-regular equipartition of order k is γ-super-regular if all but $\gamma\binom{k}{2}$ of the pairs are γ-regular with density $p \pm \gamma$ or all but $\gamma\binom{k}{2}$ of the pairs are γ-regular with density $\overline{p} \pm \gamma$. We need to show that any graph G that satisfies $\mathcal{P}^*_H$ must be either p-quasi-random or $\overline{p}$-quasi-random[7]. Fix any $\zeta > 0$ and recall that by Lemma 2.3 we know that in order to show that a graph G has the property that every set $U \subseteq V(G)$ satisfies $e(U) = \frac{1}{2}p|U^2| \pm \zeta n^2$ or that every such set satisfies $e(U) = \frac{1}{2}\overline{p}|U^2| \pm \zeta n^2$, it is enough to show that G has an ϵ-super-regular partition of order at least t, where

$$t = t_{2.3}(\zeta)\ , \tag{2}$$

and

$$\epsilon = \epsilon_{2.3}(\zeta)\ . \tag{3}$$

Let us define the following constants[8]

$$r = \max(r_{2.5}(\epsilon/2),\ N_{2.1}(h)) \tag{4}$$

$$\gamma = \min(\epsilon,\ \gamma_{2.5}(r)\ , \gamma_{2.4}(\delta_{2.1}(\epsilon, h, r)/2, h)) \tag{5}$$

$$N = \max(t,\ N_{2.6}(\gamma, r),\ N_{2.5}(r),\ N_{2.1}(h)) \tag{6}$$

$$T = T_{2.2}(\gamma/r^2,\ N) \tag{7}$$

[7]Recall that the meaning of that is that either every set $U \subseteq V(G)$ satisfies $e(U) = \frac{1}{2}p|U^2| \pm \zeta n^2$ or that every such set satisfies $e(U) = \frac{1}{2}\overline{p}|U^2| \pm \zeta n^2$ for some small $\zeta > 0$

[8]We note that we need N, which is defined in (6), in order to allow us to apply the various lemmas we stated above, that all work for large enough graphs.

$$\delta = \delta_{2.1}(\epsilon, h, r)/2T. \tag{8}$$

To complete the proof that $\mathcal{P}^*_H$ implies that a graph is either p-quasi-random or $\overline{p}$-quasi-random, we show (via Lemma 2.3) that for any $\zeta > 0$ there is an $N(\zeta)$ and $\delta(\zeta)$ such that the following holds: if G is a graph on at least $N(\zeta)$ vertices and for every h-tuple of vertex sets $U_1, \ldots, U_h \subseteq V(G)$ of (arbitrary) size m each, and for every $\sigma : [h] \to [h]$ we have

$$H^*_\sigma[U_1, \ldots, U_h] = \delta_H(p)m^h \pm \delta(\zeta)n^h \ , \tag{9}$$

then G has an ϵ-super-regular partition of order k, where $k \geq t$. We will show that one can take $N(\zeta)$ to be the integer N defined in (9) and that $\delta(\zeta)$ can be taken as the value defined in (8).

So let G be a graph of size at least N, and apply Lemma 2.2 (the regularity lemma) on G with γ/r^2 and N that were defined in (5) and (6). Lemma 2.2 guarantees that G has a γ-regular partition $\mathcal{V} = \{V_1, \ldots, V_k\}$ where $t \leq N \leq k \leq T$ and T is given in (7). We now need to show that all but $\epsilon\binom{k}{2}$ of the pairs (V_i, V_j) are ϵ-regular and satisfy $d(V_i, V_j) = p \pm \epsilon$ or ϵ-regular and satisfy $d(V_i, V_j) = \overline{p} \pm \epsilon$. Let us define W to be a weighted graph on k vertices, where if (V_i, V_j) is ϵ-regular then (i, j) are connected with an edge of weight $w(i, j) = d(V_i, V_j)$, and if (V_i, V_j) is not ϵ-regular then (i, j) are not connected. So our goal is to show that either all but $\epsilon\binom{k}{2}$ of the pairs of vertices of W are connected by an edge with weight $p \pm \epsilon$ or that all but $\epsilon\binom{k}{2}$ of the pairs of vertices of W are connected by an edge with weight $\overline{p} \pm \epsilon$.

Claim 2.7 *Any copy of K_r in W satisfies the following:*

1. *Any edge has either weight $p \pm \epsilon$ or weight $\overline{p} \pm \epsilon$.*

2. *If $p \neq \overline{p}$ then either at most $r - 1$ of them have weight $p \pm \epsilon$ or at most $r - 1$ of them have weight $\overline{p} \pm \epsilon$.*

Proof: Consider any copy of K_r in W and suppose wlog that the vertices of this copy are $1, \ldots, r$. Recall that $k \leq T$ and that G satisfies (9) with the δ that was chosen in (8). We thus infer that for any injective mapping $\phi : [h] \to [r]$

$$H^*_\phi[V_1, \ldots, V_r] = \delta_H(p)\left(\frac{n}{k}\right)^h \pm \delta n^h = (\delta_H(p) \pm \frac{1}{2}\delta_{2.1}(\epsilon, h, r))\left(\frac{n}{k}\right)^h \ . \tag{10}$$

In addition, as we are referring to r vertices that form a copy of K_r in W, we know that $V_1, \ldots, V_r$ are all pairwise γ-regular. Thus the choice of γ in (5) guarantees via Lemma 2.4 that for any injective mapping $\phi : [h] \to [r]$ we have

$$H^*_\phi[V_1, \ldots, V_r] = (W(\phi) \pm \frac{1}{2}\delta_{2.1}(\epsilon, h, r))\left(\frac{n}{k}\right)^h \ . \tag{11}$$

Combining (10) and (11) we infer that for any $\phi : [h] \to [r]$ we have $W(\phi) = \delta_H(p) \pm \delta_{2.1}(\epsilon, h, r)$. Hence, the two assertions of the claim follow from Lemma 2.1. ■

We are now going to use Lemma 2.1 in order to color some of the edges of W. Consider any copy of K_r in W. We know from the first assertion of Claim 2.7 that all the edge weights in the copy of K_r are either $p \pm \epsilon$ or $\overline{p} \pm \epsilon$. If $p = \overline{p}$ then we color all the edges of this K_r with the color red. So assume that $p \neq \overline{p}$ and recall that we assume wlog that in this case $\epsilon < |p - \overline{p}|/2$, which makes it possible to color the edges whose weight is $p \pm \epsilon$ blue, and the edges whose weight is $\overline{p} \pm \epsilon$ red (in a well defined way). We now apply this coloring scheme to any copy of K_r in W. We claim that we have thus colored at least $(1-\gamma)\binom{k}{2}$ of the edges of W. Indeed, as we applied the regularity lemma with γ/r^2 we know that W has at least $(1-\gamma/r^2)\binom{k}{2}$ edges. As $k \geq N_{2.6}(\gamma, r)$ we infer from Lemma 2.6 that at least $(1-\gamma)\binom{k}{2}$ of the edges of W belong to a copy of K_r thus they are colored in the above process. Let us now remove from W all the uncolored edges and call the new graph W'. Thus W' has at least $(1-\gamma)\binom{k}{2}$ edges and they are all colored either red or blue.

We now claim that either W' has at most $\frac{\epsilon}{2}\binom{k}{2}$ red edges, or at most $\frac{\epsilon}{2}\binom{k}{2}$ blue edges. Indeed, if W' had at least $\epsilon\binom{k}{2}$ red edges *and* at least $\epsilon\binom{k}{2}$ blue edges, then our choice of r and γ in (4) and (5), the fact that W' has at least $(1-\gamma)\binom{k}{2}$ edges, and that $k \geq N_{2.5}(r)$, would allow us to apply Lemma 2.5 on W' and infer that it has a copy of K_r with at least r blue edges and at least r red edges, contradicting Claim 2.7 (recall that W' is a subgraph of W).

We thus conclude that W has at least $(1-\gamma)\binom{k}{2} \geq (1-\frac{\epsilon}{2})\binom{k}{2}$ edges, and that even if $p \neq \overline{p}$ either all but $\frac{\epsilon}{2}\binom{k}{2}$ of them are red or all but at most $\frac{\epsilon}{2}\binom{k}{2}$ of them are blue. By the definition of W, this means that in the equipartition $\mathcal{V}$ either all but $\epsilon\binom{k}{2}$ of the pairs are ϵ-regular with density $p \pm \epsilon$, or all but at most $\epsilon\binom{k}{2}$ of them are ϵ-regular with density $\overline{p} \pm \epsilon$, which completes the proof. ■

2.4 Proofs of additional lemmas

We end this section with the proofs of Lemmas 2.3, 2.5 and 2.6.

Proof of Lemma 2.3: We claim that one can take $\epsilon = \epsilon_{2.3}(\zeta) = \frac{1}{8}\zeta$ and $t = t_{2.3}(\zeta) = 8/\zeta$. Indeed, suppose G has an ϵ-super-regular partition $\{V_1, \ldots, V_k\}$ of order $k \geq t$, that is, a partition in which all but $\epsilon\binom{k}{2}$ of the pairs (V_i, V_j) are ϵ-regular with density $d(V_i, V_j) = p \pm \epsilon$. Let us count the number of edges of G that do not connect a pair (V_i, V_j) which is ϵ-regular with density $d(V_i, V_j) = p \pm \epsilon$. As $k \geq t \geq 8/\zeta$ we know that the number of pairs of vertices that both belong to the same set V_i is at most $k|n/k|^2 \leq \frac{1}{8}\zeta n^2$. As all but $\epsilon\binom{k}{2}$ of the pairs (V_i, V_j) are ϵ-regular with density $d(V_i, V_j) = p \pm \epsilon$, we also know that the number of pairs connecting pairs (V_i, V_j), which do not satisfy these two conditions, is bounded by $\epsilon\binom{k}{2}(n/k)^2 \leq \frac{1}{8}\zeta n^2$.

Consider now a set of vertices U, and define $U_i = U \cap V_i$. The number of vertices of U that belong to a set U_i whose size is smaller than $\epsilon|V_i|$ is bounded by ϵn. Therefore the number of pairs of vertices of U such that one of them belongs to a set U_i of size smaller than $\epsilon|V_i|$ is bounded by $\epsilon n^2 \leq \frac{1}{8}\zeta n^2$. Combining the above three facts we conclude that all by $\frac{1}{2}\zeta n^2$ of the pairs of vertices of $u_i, u_j \in U$ are such that: (1) $u_i \in U_i$, $u_j \in U_j$ and $i \neq j$; (2) $U_i \geq \epsilon|V_i|$ and $U_j \geq \epsilon|V_j|$; (3) (V_i, V_j) is ϵ-regular with density $p \pm \epsilon$. Therefore, by the definition of a regular pair we get that the density of U in all but $\frac{1}{2}\zeta n^2$ of its pairs is $p \pm 2\epsilon = p \pm \frac{1}{2}\zeta$, and therefore $e(U) = \frac{1}{2}p \pm \zeta n^2$. ■

Proof of Lemma 2.5: Suppose we randomly pick r vertices $v_1, \ldots, v_r$ from K *with repetitions* where $r = \Omega(1/\epsilon^3)$. Clearly, if $k \geq 10r^2$ then by a Birthday-Paradox argument we infer that with probability at least 3/4 all the vertices $v_1, \ldots, v_r$ are distinct. Suppose wlog that r is even and let us partition the set of unordered pairs (v_i, v_j) into $r-1$ perfect matchings $M_1, \ldots, M_{r-1}$ on the vertices $v_1, \ldots, v_r$. For every pair i, j let $p_{i,j}$ be the indicator random variable for the event that r_i

and r_j are connected in K by a red edge. As we sample with repetitions then for every matching M_t, the $r/2$ events $\{p_{i,j} : (i,j) \in M_t\}$ are independent. Also, as K has at least $\epsilon\binom{k}{2}$ red edges, we have that $Pr[p_{i,j} = 1] \geq \epsilon/2$ (we lose a little due to the probability of having non distinct vertices). We thus conclude that for any matching M_t, the expected number of red edges spanned by the pairs $\{(i,j) : (i,j) \in M_t\}$ is at least $\epsilon r/4$ and by a Chernoff bound, the probability of deviating from this expectation by more than $\epsilon r/8$ is bounded by $2^{-\Theta(\epsilon^2 r)} < 1/4r$. Clearly the same analysis applies for the red edges. We conclude by the union bound that with probability at least 3/4 the r vertices span at least $\epsilon r^2/16$ red edges and at least $\epsilon r^2/16$ blue edges. As $r = \Omega(1/\epsilon^3)$ we have $\epsilon r^2/16 \geq r$ therefore we have the required amount of red/blue edges. We conclude that one can take $r_{2.5} = \Omega(1/\epsilon^3)$ and $N_{2.5}(r) = 10r^2$.

Finally, to conclude that all the pairs (r_i, r_j) are connected we take $\gamma_{2.5} = 1/4r^2$. This way, the probability that a pair of vertices are not connected is at most $1/4r^2$ and by the union bound, with probability at least 3/4 they are all connected. So to recap, if we sample with repetition r vertices, then with probability at least 1/4 they are all distinct, all connected, and have at least r red edges and at least r blue edges. So there must be at least one such set of r vertices in K. ■

Proof of Lemma 2.6: Suppose k is large enough to guarantee by Rödl's theorem [18] that the complete graph on k vertices contains $(1-\gamma)\binom{k}{2}/\binom{r}{2}$ edge disjoint copies of K_r. If we now consider the same copies of K_r in K (more precisely, the vertex sets of these copies) then the fact that K has $(1-\gamma)\binom{k}{2}$ edges implies that at most $\gamma\binom{k}{2} = \gamma\binom{r}{2} \cdot \binom{k}{2}/\binom{r}{2}$ of these copies of K_r have a pair of vertices that are not connected. Thus, K contains at least $(1-\gamma r^2)\binom{k}{2}/\binom{r}{2}$ edge disjoint copies of K_r implying that at least $(1-\gamma r^2)\binom{k}{2}$ edges of K belong to a copy of K_r. ■

3 Proof of the Key Lemma

As in Section 2, let us fix a real $0 < p < 1$ and a fixed graph H on h vertices. Let also $\overline{p}$ be the conjugate of p with respect to H. We will again work with weighted complete graphs W on r vertices, and will identify the vertices of W with $[r]$ and the vertices of H with $[h]$. Each pair of vertices $1 \leq i < j \leq r$ of W has a weight $0 \leq w(i,j) \leq 1$ that is given by some weight function $w : E(W) \to [0,1]$. We remind the reader of the notation $W(\phi)$ that was introduced at the beginning of Section 2.

Recall that Lemma 2.1 states that if all the values $W(\phi)$ are close to what they should be, then all the weights $w(i,j)$ are close to what they should be. The following lemma is an "exact" version of Lemma 2.1 where we assume that the values $W(\phi)$ are exactly what they should be. The proof of Lemma 2.1 will then follow from the lemma below using standard continuity arguments. Observe that the lemma below actually gives a bit more information than what we need for Lemma 2.1. In what follows let Φ be the set of all possible injective mappings $\phi : [h] \to [r]$, and notice that there are $r!/(r-h)!$ elements in Φ.

Lemma 3.1 *For every $h > 2$ there exists $N_{3.1} = N_{3.1}(h)$ so that the following holds. Let H be a fixed graph with m edges and h vertices. If $r \geq N_{3.1}$ and W is a labeled weighted graph on r vertices satisfying $W(\phi) = \delta_H(p)$ for all $\phi \in \Phi$, then $w(i,j) \in \{p, \overline{p}\}$ for all $1 \leq i < j \leq r$. Furthermore,* *if* $\gcd(\binom{h}{2}, m) = 1$ *then all edge weights are the same, and if* $\gcd(\binom{h}{2}, m) > 1$ *then either all edge weights are the same, or else there exists one vertex whose deletion from* W *yields a subgraph with* $r-1$ *vertices all of whose edge weights are the same.*

We split the proof of Lemma 3.1 into two parts. We initially prove Lemma 3.2 below showing that all $w(i,j)$ are taken from $\{p, \overline{p}\}$. We then use this lemma in order to show that in fact most of the $w(i,j)$ are either p or $\overline{p}$.

An important ingredient in the proof of Lemma 3.2 will be a theorem of Gottlieb [3], concerning the rank of set inclusion matrices. For integers $r \geq h > 2$, the *inclusion matrix* $A(r,h)$ is defined as follows: The rows of $A(r,h)$ are indexed by h-element subsets of $[r]$, and the columns by the 2-element subsets of $[r]$. Entry (i,j) of $A(r,h)$ is 1 if the 2-element set, whose index is j, is contained in the r-element set, whose index is i. Otherwise, this entry is 0. Notice that $A(r,h)$ is a square matrix if and only if $r = h+2$, and that for $r > h+2$, $A(r,h)$ has more rows than columns. Trivially, $rank(A(r,h)) \leq \binom{r}{2}$. However, Gottlieb [3] proved [9] that in fact

Theorem 4 (Gottlieb [3]) $rank(A(r,h)) = \binom{r}{2}$ *for all* $r \geq h+2$.

Lemma 3.2 *Let* H *be a fixed graph with* $h > 2$ *vertices. If* $r \geq h+2$ *and* W *is a labeled weighted graph on* r *vertices satisfying* $W(\phi) = \delta_H(p)$ *for all* $\phi \in \Phi$*, then* $w(i,j) \in \{p, \overline{p}\}$ *for all* $1 \leq i < j \leq r$.

Proof: We associate a variable $x_{i,j}$ for each $1 \leq i < j \leq r$, which represents the unknown $w(i,j)$. Thus, for any $\phi \in \Phi$ we have that $W(\phi)$ is given by the polynomial

$$P_\phi = \prod_{(i,j) \in E(H)} x_{\phi(i),\phi(j)} \prod_{(i,j) \notin E(H)} (1 - x_{\phi(i),\phi(j)}) \ . \tag{12}$$

As our assumption is that $W(\phi) = \delta_H(p)$ for all $\phi \in \Phi$ we have the following set of $r!/(r-h)!$ polynomial equations E_ϕ:

$$E_\phi : \qquad \prod_{(i,j) \in E(H)} x_{\phi(i),\phi(j)} \prod_{(i,j) \notin E(H)} (1 - x_{\phi(i),\phi(j)}) = \delta_H(p) \ . \tag{13}$$

Our goal now is to show that the only solution to this system is $x_{i,j} \in \{p, \overline{p}\}$.

For a vertex set $S \subseteq [r]$ of size h, let $E[S]$ denote the $\binom{h}{2}$ edges of W induced by S. Let $\mathcal{S}$ be the set of all h-element subsets of $V(W) = [r]$ and notice that $|\mathcal{S}| = \binom{r}{h}$. For every $S \in \mathcal{S}$ let Φ_S be the set of all $h!$ elements of Φ that are bijections on S. For every set S let us take the product of the $h!$ equations $\{E_\phi : \phi \in \Phi_S\}$ of (13). We thus get the following system of $\binom{r}{h}$ polynomial equations (one for every $S \in \mathcal{S}$) with $\binom{r}{2}$ variables (one for each $(i,j) \in E(W)$):

$$E_S : \qquad \prod_{(i,j) \in E[S]} \left(x_{i,j}^m \cdot (1 - x_{i,j})^{\binom{h}{2} - m} \right)^{2!(h-2)!} = (\delta_H(p))^{h!} \ . \tag{14}$$

[9] Gottliebs's theorem actually deals with the more general case where the columns are indexed by the d element subsets of $[r]$ where $2 \leq d \leq h$, and in that case the rank is $\binom{r}{d}$ for all $r \geq h+d$.

In order to show that the only solution of the equations E_S is given by $x_{i,j} \in \{p, \overline{p}\}$, it would be convenient to first transform them to linear equalities, by taking logarithms on both sides. Define

$$y_{i,j} = m \cdot \log(x_{i,j}) + \left(\binom{h}{2} - m\right) \cdot \log(1 - x_{i,j}) \tag{15}$$

and note that if we take logarithm of the equations given in (14) and use the $y_{i,j}$ defined above, we thus obtain an equivalent system of *linear* equations on the $\binom{r}{2}$ variables $y_{i,j}$, where equation E_S becomes

$$E'_S : \qquad \sum_{(i,j)\in E[S]} y_{i,j} = \binom{h}{2} \cdot \log(\delta_H(p)) \ . \tag{16}$$

We can write the $\binom{r}{h}$ linear equations E'_S as $Ax = b$ where A in an $\binom{r}{h} \times \binom{r}{2}$ matrix, and b is the all $\binom{h}{2} \cdot \log(\delta_H(p))$ vector. A key observation at this point is that A is precisely the inclusion matrix $A(r, h)$. Since $r \geq h + 2$ we obtain, by Theorem 4, that the system has a unique solution and the values of the variables $y_{i,j}$ are uniquely determined. Now, as each set $S \in \mathcal{S}$ is of size h it is clear that setting $y_{i,j} = \log(\delta_H(p))$ for all (i, j) gives a valid solution of the linear equations given in (16), and by the above observation, this is in fact the unique solution. Recalling the definition of $y_{i,j}$ in (15), this implies that for all (i, j) we have

$$x_{i,j}^m \cdot (1 - x_{i,j})^{\binom{h}{2}-m} = \delta_H(p) \ .$$

Now, as $\{p, \overline{p}\}$ are the only solutions to the above equation, we deduce that indeed $x_{i,j} \in \{p, \overline{p}\}$, proving the lemma. ■

For the proof of Lemma 3.1, we will need another simple lemma. A graph is called *pairwise regular* if there exists a number t so that $d(x) + d(y) - d(x, y) = t$ for all pairs of distinct vertices x, y. Here $d(v)$ denotes the degree of v and $d(u, v) = 1$ if (u, v) is an edge, otherwise $d(u, v) = 0$. A graph is called *pairwise outer-regular* if there exists a number t so that $d(x) + d(y) - 2d(x, y) = t$ for all pairs of distinct vertices x, y. Trivially, a graph is pairwise regular if and only if its complement is. The same holds for pairwise outer-regular. It is also trivial that the complete graph (and the empty graph) is both pairwise regular and pairwise outer-regular. Notice, that $K_{1,2}$ is also pairwise regular, and that $K_{1,3}$ is also pairwise outer-regular. The following lemma, whose proof is deferred to the end of this section, establishes that these are the only non-trivial cases.

Lemma 3.3 *The only non-complete and non-empty graphs which are pairwise regular are $K_{1,2}$ and its complement. The only non-complete and non-empty graphs which are pairwise outer-regular are $K_{1,3}$ and its complement.*

Proof of Lemma 3.1: Notice that if $p = \overline{p}$ then there is actually nothing to prove, since Lemma 3.2 already yields the desired conclusion. Hence, assume $p \neq \overline{p}$. Observe, that this implies that H in not the complete graph nor the empty graph as in these two cases $p = \overline{p}$.

By Lemma 3.2, each edge weight is either p or $\overline{p}$. We color the edges of W with two colors: blue for edges whose weight is p and red for edges whose weight is $\overline{p}$. We may assume that our coloring is non-trivial, that is, that we have both red and blue edges, since otherwise there is nothing to prove. Each $\phi \in \Phi$ defines a labeled copy of H in W. Let $b(\phi)$ be the number of edges of H mapped to blue edges and let $a(\phi)$ be the number of non-edges[10] of H mapped to blue edges. Then, the number of edges of H mapped to red edges is $m - b(\phi)$ and the number of non-edges of H mapped to red edges is $\binom{h}{2} - m - a(\phi)$. Thus, we have for every $\phi \in \Phi$ that

$$\delta_H(p) = W(\phi) = p^{b(\phi)}\overline{p}^{m-b(\phi)}(1-p)^{a(\phi)}(1-\overline{p})^{\binom{h}{2}-m-a(\phi)} . \tag{17}$$

Multiplying (17) by $\overline{p}^{b(\phi)}(1-\overline{p})^{a(\phi)}$ we get that

$$\delta_H(p) \cdot \overline{p}^{b(\phi)}(1-\overline{p})^{a(\phi)} = p^{b(\phi)}(1-p)^{a(\phi)}\overline{p}^{m}(1-\overline{p})^{\binom{h}{2}-m} = \delta_H(p) \cdot p^{b(\phi)}(1-p)^{a(\phi)} \tag{18}$$

where in the second equality we use the fact that $\overline{p}^m(1-\overline{p})^{\binom{h}{2}-m} = \delta_H(p)$. This implies that

$$\left(\frac{p}{\overline{p}}\right)^{b(\phi)} \left(\frac{1-p}{1-\overline{p}}\right)^{a(\phi)} = 1 . \tag{19}$$

On the other hand, since p and $\overline{p}$ are both solutions of the equation $x^m \cdot (1-x)^{\binom{h}{2}-m} = \delta_H(p)$ we also know that

$$\left(\frac{p}{\overline{p}}\right)^{m} \left(\frac{1-p}{1-\overline{p}}\right)^{\binom{h}{2}-m} = 1 . \tag{20}$$

Thus, solving (19) for $\frac{p}{\overline{p}}$ and plugging it into (20) gives that for any ϕ

$$a(\phi) \cdot m = b(\phi) \cdot \left(\binom{h}{2} - m\right) . \tag{21}$$

Consider first the case where $\gcd(m, \binom{h}{2}) = 1$. This implies that $\gcd(m, \binom{h}{2} - m) = 1$. Since the red-blue coloring is not trivial there is a K_h subgraph of W which contains both red and blue edges. Thus there exists $\phi \in \Phi$ so that $0 < a(\phi) + b(\phi) < \binom{h}{2}$. There are two ways in which (21) can be satisfied: the first is if $a(\phi) = b(\phi) = 0$, but this violates the fact that $0 < a(\phi) + b(\phi)$. The second is if $a(\phi)$ is a multiple of $\binom{h}{2} - m$ and $b(\phi)$ is a multiple of m, but this violates $a(\phi) + b(\phi) < \binom{h}{2}$. Thus, the coloring must be trivial, and we are done.

Now consider the case $\gcd(m, \binom{h}{2}) > 1$. By Ramsey's Theorem if $N_{3.1}(h)$ is sufficiently large, there is a monochromatic copy of K_{3h-8} in W. Let T denote a maximal monochromatic copy in W. Thus, T has t vertices and $r > t \geq 3h - 8$. Suppose, w.l.o.g., that T is completely red. Let x be a vertex outside T. By maximality of T, there exists $y \in T$ so that (x, y) is blue. Suppose x has at least $h - 2$ red neighbors in T, say $(x, v_1), \ldots, (x, v_{h-2})$ are all red. Then, $\{x, y, v_1, \ldots, v_{h-2}\}$ induce a copy of K_h which has precisely one blue edge. If ϕ is any bijection onto this copy then $a(\phi) + b(\phi) = 1$, but this must violate (21) and hence the coloring must be trivial and we are done.

We may now assume that each vertex x outside T has at most $h - 3$ red neighbors in T. Now, if $t = r - 1$ then there are at most $r - 1$ blue edges in our coloring, all incident with x, and we are done. Otherwise, there are at least two vertices x_1 and x_2 outside T, that have at least $t - 2(h - 3) \geq 3h - 8 - 2h + 6 = h - 2$ common neighbors $\{v_1, \ldots, v_{h-2}\}$ in T so that all edges (x_i, v_j) are blue for $i = 1, 2$ and $j = 1, \ldots, h - 2$.

[10] The non-edges of H are all the pairs i, j that are not connected in H.

Consider first the case where (x_1, x_2) is blue. Since $\gcd(m, \binom{h}{2}) > 1$ we must have that H is not $K_{1,2}$ nor its complement. Thus, by Lemma 3.3, H is not pairwise regular[11]. Let $\{u_1, u_2\}$ and $\{u_3, u_4\}$ be two pairs of distinct vertices of H so that

$$d(u_1) + d(u_2) - d(u_1, u_2) \neq d(u_3) + d(u_4) - d(u_3, u_4) \; . \tag{22}$$

Let ϕ_1 be a bijection from $V(H)$ to $\{x_1, x_2, v_1, \ldots, v_{h-2}\}$ mapping u_1 to x_1 and u_2 to x_2. Clearly, $b(\phi_1) = d(u_1) + d(u_2) - d(u_1, u_2)$. Similarly, if ϕ_2 is a bijection from $V(H)$ to $\{x_1, x_2, v_1, \ldots, v_{h-2}\}$ mapping u_3 to x_1 and u_4 to x_2 then $b(\phi_2) = d(u_3) + d(u_4) - d(u_3, u_4)$. In particular, we get from (22) that $b(\phi_1) \neq b(\phi_2)$. We claim however that this is impossible as in fact $b(\phi_1) = b(\phi_2)$. Indeed, by combining (21) for ϕ_1 and for ϕ_2 we get that $a(\phi_1)/a(\phi_2) = b(\phi_1)/b(\phi_2)$. Further we have $b(\phi_1) + a(\phi_1) = b(\phi_2) + a(\phi_2)$ as both sides are equal to the number of blue edges in the corresponding induced K_h of W. Combining the two equations we get $b(\phi_1) = b(\phi_2)$.

Consider finally the case where (x_1, x_2) is red. Assume first that H is not $K_{1,3}$ nor its complement. Thus, by Lemma 3.3, H is not pairwise outer-regular. Let $\{u_1, u_2\}$ and $\{u_3, u_4\}$ be two pairs of distinct vertices of H so that

$$d(u_1) + d(u_2) - 2d(u_1, u_2) \neq d(u_3) + d(u_4) - 2d(u_3, u_4) \; .$$

Let ϕ_1 be a bijection from $V(H)$ to $\{x_1, x_2, v_1, \ldots, v_{h-2}\}$ mapping u_1 to x_1 and u_2 to x_2. Clearly, $b(\phi_1) = d(u_1) + d(u_2) - 2d(u_1, u_2)$. Similarly, if ϕ_2 is a bijection from $V(H)$ to $\{x_1, x_2, v_1, \ldots, v_{h-2}\}$ mapping u_3 to x_1 and u_4 to x_2 then $b(\phi_2) = d(u_3) + d(u_4) - 2d(u_3, u_4)$. In particular, $b(\phi_1) \neq b(\phi_2)$. As in the previous case, this is a contradiction. If $H = K_{1,3}$ then $h = 4$ and we can use the fact that x_1 has at least $3h - 8 - (h-3) = 3$ blue neighbors in T denoted y_1, y_2, y_3. Thus, x_1, y_1, y_2, y_3 have a red triangle and a blue $K_{1,3}$. Let ϕ_1 map the vertex of degree 3 of H to x_1 and the rest to y_1, y_2, y_3, yielding $b(\phi_1) = 3$. Let ϕ_2 map the vertex of degree 3 of H to y_1 and the rest to x_1, y_2, y_3 yielding $b(\phi_2) = 1$. Again, $b(\phi_1) \neq b(\phi_2)$, a contradiction. The case of the complement of $K_{1,3}$ is proved in the same way. ■

For the proof of Lemma 2.1, we will need the following simple fact

Claim 3.4 *For any integer p and δ there is a $\gamma = \gamma_{3.4}(\delta, p)$ with the following property: Let A be any $p \times p$ non-singular $0/1$ matrix, let b be any vector in $\mathbb{R}^p$ and let $x \in \mathbb{R}^p$ be the unique solution of the system of linear equations $Ax = b$. Then if b' satisfies $\ell_\infty(b', b) \leq \gamma$ then the unique solution x' of $Ax' = b'$ satisfies $\ell_\infty(x', x) \leq \gamma$.*

Proof: Fix any $p \times p$ non-singular matrix A with $0/1$ entries. Then the solution of $Ax = b$ is given by $x = A^{-1}b$. As $x_i = \sum_{j=1}^{p} A^{-1}_{i,j} \cdot b_j$ is a continuous function of b it is clear to for any δ there is a $\gamma = \gamma(\delta, A)$ such that if $\ell_\infty(b', b) \leq \gamma$ then the unique solution x' of $Ax' = b'$ satisfies $\ell_\infty(x', x) \leq \delta$. Now, as there are finitely many $0/1$ $p \times p$ matrices, we can set $\gamma = \gamma_{3.4}(\delta, p) = \min_A \gamma(\delta, A)$, where the minimum is taken over all $0/1$ $p \times p$ matrices. ■

[11] Remember that at this point we know that H is neither a complete graph nor an edgeless graph.

Proof of Lemma 2.1: The lemma is an immediate consequence of Lemma 3.1 using standard arguments of continuity; the continuity of polynomials as functions, and the continuity of unique solutions to linear systems that is given in Lemma 3.4 above. First we can take $N_{2.1}(h) = N_{3.1}(h)$. Now, given any $r \geq N_{3.1}(h)$ and ϵ we need to show that if all $W(\phi)$ are very close to $\delta_H(p)$ then we can get the conclusion of Lemma 2.1.

First, we see that in Lemma 3.2 if all $W(\phi)$ are close to $\delta_H(p)$ then by Lemma 3.4 any solution to the linear equations E'_S given in (16) satisfies that all $y_{i,j}$ are very close to $\log(\delta_H(p))$. By continuity of 2^x this means that $x_{i,j}^m(1-x_{i,j})^{\binom{h}{2}-m}$ is close to $\delta_H(p)$, which again by continuity of x^k implies that either $x_{i,j}$ is close to p or to $\overline{p}$. So the conclusion of Lemma 3.2 is that if all $W(\phi)$ are close to $\delta_H(p)$, then all densities are indeed close to either p or $\overline{p}$.

For the rest of the proof, in equations (17) and (18) we replace p and $\overline{p}$ with quantities close to them. This means that (19) and (20) are no longer equations but approximately equal to 1. This implies that in (21) we also have approximate equality. However, note that as both sides of (21) involve integers, once the two sides are close enough, they must in fact be equal. Now, as the rest of the proof only relies on the validity of (21) it follows verbatim as in the proof of Lemma 3.1. ■

It is interesting to note that we cannot hope to prove a stronger version of Lemma 3.1 in which *all* edge weights are the same, regardless of $\gcd(\binom{h}{2}, m)$. Indeed, consider the case where $H = C_h$ is a cycle with $h \geq 4$ vertices. For every $r \geq h+1$, there are weighted complete graphs W with r vertices having $W(\phi) = \delta_H(p)$ for each $\phi \in \Phi$, while still some edges of W have weight p and others have weight $\overline{p}$. Indeed, assume that all weights of edges not incident with $r \in W$ have weight p, and the $r-1$ edges incident with r has weight $\overline{p}$. Now, if the image of ϕ does not contain r then, clearly,

$$W(\phi) = p^h(1-p)^{\binom{h}{2}-h} = \delta_H(p).$$

On the other hand, if the image of ϕ contains r then

$$W(\phi) = p^{h-2}\overline{p}^2(1-p)^{\binom{h}{2}-2h+3}(1-\overline{p})^{h-3}.$$

But not that $\overline{p}^2(1-\overline{p})^{h-3}$ is just $\delta_H(p)^{2/h}$, and hence it also equals $p^2(1-p)^{h-3}$. Consequently, $W(\phi) = \delta_H(p)$ in this case as well.

Proof of Lemma 3.3: Let us say that $(x_1, y_1)(x_2, y_2)$ are violating, with respect to the property of being pairwise regular if $d(x_1) + d(y_1) - d(x_1, y_1) \neq d(x_2) + d(y_2) - d(x_2, y_2)$ and violating with respect to the property of being pairwise outer-regular if $d(x_1)+d(y_1)-2d(x_1, y_1) \neq d(x_2)+d(y_2)-2d(x_2, y_2)$. Suppose first that G is a pairwise regular graph which is neither complete nor empty. We claim that this implies that $|d(x) - d(y)| \leq 1$ for any two vertices $x, y \in V(G)$. Indeed, if there is a pair that violates this, then $(x, z), (y, z)$ is violating for any z. Note that G cannot be regular, otherwise $(x_1, y_1)(x_2, y_2)$ is violating whenever (x_1, y_1) is and edge and (x_2, y_2) is not. So partition the vertices of G into two non empty sets, V_1 and V_2, where all the vertices of V_1 have degree s and those of V_2 have degree $s-1$.

If $|V_1| > 1$ then V_1 must be a clique otherwise $(x_1, y)(x_2, y)$ is violating for any $x_1, x_2 \in V_1$ and $y \in V_2$. In particular, we have $t = 2s - 1$. We also have that $|V_2| = 1$ as otherwise $(x_1, x_2)(y_1, y_2)$ is violating for any $x_1, x_2 \in V_1$ and $y_1, y_2 \in V_2$. If the unique vertex v of V_2 is connected to $x_1 \in V_1$

but not to $x_2 \in V_1$ then $(x_1, v)(x_2, v)$ is clearly violating, so v is either connected to all the vertices of V_1 or else is an isolated vertex. If v is an isolated vertex then $s = 1$, which implies that V_1 is a clique of size 2, and G is thus the complement of $K_{1,2}$. If v is connected to all the vertices of V_1 then $|V_1| = s - 1$ which is impossible, since in a graph with s vertices there cannot be vertices with degree s. If $|V_1| = 1$ then we must have $|V_2| > 1$. Note that in this case V_2 must span an independent set as otherwise $(x, y_1)(y_1, y_2)$ is violating for any choice of $y_1, y_2 \in V_2$ and $x \in V_1$. As G is not edgeless we infer that $s - 1 = 1$ implying that G is $K_{1,2}$.

Suppose now that G is pairwise outer-regular and is neither complete nor empty. Following the same reasoning as above, we must have for any two vertices x, y of G, that $|d(x) - d(y)| \leq 2$. Again, note that G cannot be regular, so partition the vertices of G into two non-empty sets, V_1 and V_2, where all of the vertices of V_1 have degree s and all the vertices of V_2 have degree $s - 1$ or $s - 2$. If $|V_1| > 1$ then again V_1 must span a clique, as otherwise $(x_1, y)(x_2, y)$ is violating for any $x_1, x_2 \in V_1$ and $y \in V_2$, and therefore $t = 2s - 2$. Note that if $y \in V_2$ is connected to $x \in V_1$ then $(x_1, x)(x, y)$ is violating for any other $x_1 \in V_1$. Also, if $|V_2| \geq 2$ then any pair of vertices of V_2 must be disconnected with degree $s - 1$ as otherwise $(x_1, x_2)(y_1, y_2)$ is violating for any $x_1, x_2 \in V_1$ and $y_1, y_2 \in V_2$. We thus get that the degree of vertices in V_2 is zero, hence either $s = 1$ or $s = 2$. The former case implies that $t = 0$ and that $|V_1| = 2$. This means that G has just one edge, which is not an outer-regular graph. If $s = 2$ then $|V_1| = 3$ implying that G is a triangle plus some isolated vertices. If there is one such vertex then G is the complement of $K_{1,3}$, and if there are two such vertices y_1, y_2, then $(y_1, y_2)(y_1, v)$ is violating for any $v \in V_1$. So assume that $|V_1| = 1$, which implies that $|V_2| \geq 2$. Let x be the unique vertex of V_1, and observe that if x is connected to $y_1 \in V_2$ but not to $y_2 \in V_2$ then $(x, y_1)(x, y_2)$ is violating. So either v is connected to all the vertices of V_2 or to none of them, but note that the latter case is impossible as $s > s - 1 \geq 0$. We now claim that V_2 must be edgeless. Indeed if $y_1, y_2 \in V_2$ are connected and $d(y_1) \geq d(y_2)$ then $(y_1, y_2)(x, y_1)$ is violating. We infer that the degree of the vertices of V_2 is 1, so G is either $K_{1,2}$, which is not outer regular, or $K_{1,3}$. ∎

References

[1] N. Alon, R. A. Duke, H. Lefmann, V. Rödl and R. Yuster, The algorithmic aspects of the regularity lemma, J. Algorithms 16 (1994), 80-109.

[2] D. de Caen, A note on the ranks of set-inclusion matrices, Elect. J. Combin. 8 (2001) No 5.

[3] D. H. Gottlieb, A class of incidence matrices, Proc. Amer. Math. Soc. 17 (1966), 1233-1237.

[4] F. R. K. Chung, Regularity lemmas for hypergraphs and quasi-randomness, Random Structures and Algorithms 2 (1991), 241-252.

[5] F. R. K. Chung and R. L. Graham, Quasi-random set systems, Journal of the AMS, 4 (1991), 151-196.

[6] F. R. K. Chung and R. L. Graham, Quasi-random tournaments, J. Graph Theory 15 (1991), 173-198.

[7] F. R. K. Chung and R. L. Graham, Quasi-random hypergraphs, Random Structures and Algorithms 1 (1990), 105-124.

[8] F. R. K. Chung and R. L. Graham, Maximum cuts and quasi-random graphs, Random Graphs, (Poznan Conf., 1989) Wiley-Intersci, Publ. vol 2, 23-33.

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

[10] E. Fischer, The difficulty of testing for isomorphism against a graph that is given in advance, SIAM Journal on Computing, 34 (2005), 1147-1158.

[11] T. Gowers, Quasirandom groups, manuscript, 2006.

[12] T. Gowers, Quasirandomness, counting and regularity for 3-uniform hypergraphs, manuscript, 2006.

[13] T. Gowers, Hypergraph regularity and the multidimensional Szemerédi theorem, manuscript, 2006.

[14] S. Hoory, N. Linial and A. Wigderson, Expander graphs and their applications, Bulletin of the AMS, Vol 43 (4), 2006, 439-561.

[15] M Krivelevich and B. Sudakov, Pseudo-random graphs, More sets, graphs and numbers, E. Gyori, G. O. H. Katona and L. Lovász, Eds., Bolyai Society Mathematical Studies Vol. 15, 199-262.

[16] L. Lovász and V. T. Sós, Generalized quasirandom graphs, manuscript, 2005.

[17] B. Nagle, V. Rödl and M. Schacht, The counting lemma for regular k-uniform hypergraphs, Random Structures and Algorithms 28 (2006), 113-179.

[18] V. Rödl, On a packing and covering problem, Europ. J. Combin. 6 (1985), 69–78.

[19] V. Rödl and J. Skokan, Regularity lemma for k-uniform hypergraphs, Random Structures and Algorithms 25 (2004), 1-42.

[20] A. Shapira, Quasi-randomness and the distribution of copies of a fixed graph, manuscript, 2007.

[21] M. Simonovits and V. T. Sós, Szemerédi's partition and quasirandomness, Random structures and algorithms, 2 (1991), 1-10.

[22] M. Simonovits and V. T. Sós, Hereditarily extended properties, quasi-random graphs and not necessarily induced subgraphs, Combinatorica 17 (1997), 577-596.

[23] M. Simonovits and V. T. Sós, Hereditarily extended properties, quasi-random graphs and induced subgraphs, Combinatorics Probability and Computing, 12 (2003), 319-344.

[24] E. Szemerédi, Integer sets containing no k elements in arithmetic progression, Acta Arith. 27 (1975), 299-345.

[25] E. Szemerédi, Regular partitions of graphs, In: *Proc. Colloque Inter. CNRS* (J. C. Bermond, J. C. Fournier, M. Las Vergnas and D. Sotteau, eds.), 1978, 399–401.

[26] T. Tao, A variant of the hypergraph removal lemma, J. Combin. Theory, Ser. A 113 (2006), 1257-1280.

[27] A. Thomason, Pseudo-random graphs, Proc. of Random Graphs, Poznań 1985, M. Karoński, ed., Annals of Discrete Math. 33 (North Holland 1987), 307-331.

[28] A. Thomason, Random graphs, strongly regular graphs and pseudo-random graphs, Surveys in Combinatorics, C. Whitehead, ed., LMS Lecture Note Series 123 (1987), 173-195.

[29] R. M. Wilson, Decomposition of complete graphs into subgraphs isomorphic to a given graph, Congressus Numerantium XV (1975), 647-659.

[30] R. M. Wilson, A diagonal form for the incidence matrix of t-subsets vs. k-subsets, Europ. J. Combin. 11 (1990), 609-615.