Quasi-Randomness and the Distribution of Copies of a Fixed Graph

Asaf Shapira *

Abstract

We show that if a graph G has the property that all subsets of vertices of size $n/4$ contain the "correct" number of triangles one would expect to find in a random graph $G(n, \frac{1}{2})$, then G behaves like a random graph, that is, it is quasi-random in the sense of Chung, Graham, and Wilson [6]. This answers positively an open problem of Simonovits and Sós [10], who showed that in order to deduce that G is quasi-random one needs to assume that all sets of vertices have the correct number of triangles. A similar improvement of [10] is also obtained for any fixed graph other than the triangle, and for any edge density other than $\frac{1}{2}$. The proof relies on a theorem of Gottlieb [7] in algebraic combinatorics, concerning the rank of set inclusion matrices.

1 Introduction

The theory of quasi-random graphs deals with the following fundamental question: which properties $\mathcal{P}$ of graphs have the property, that any graph that satisfies $\mathcal{P}$ "behaves" like an appropriate random graph. When we say behaves, we mean that it satisfies the properties that a random graph would satisfy with high probability. Such properties are called *quasi-random*. The study of quasi-random graphs was initiated by Thomason [11, 12] and then followed by Chung, Graham and Wilson [6]. Following the results on quasi-random graphs, quasi-random properties were also studied in various other contexts such as set systems [2], tournaments [3] and hypergraphs [4]. There are also some very recent results on quasi-random groups [8] and generalized quasi-random graphs [9].

Let us state the fundamental theorem of quasi-random graphs. A labeled copy of a graph H in a graph G is an injective mapping $\phi : V(H) \to V(G)$, that maps edges to edges, that is $(i, j) \in E(H) \Rightarrow (\phi(i), \phi(j)) \in E(G)$. For a set of vertices $U \subseteq V$ we denote by $H[U]$ the number of labeled copies of H spanned by U, and by $e(U)$ the number of edges spanned by U. A graph sequence (G_n) is an infinite set of graphs $\{G_1, G_2, \ldots\}$ of increasing size, where we denote by G_n the n^{th} graph in the sequence whose size is n. The following is (part of [1]) the main result of [6]:

Theorem 1 (Chung, Graham and Wilson [6]) *Fix any $0 < p < 1$. For any graph sequence (G_n) the following properties are equivalent:*

$\mathcal{P}_1(t)$: *For an even $t \geq 4$, let C_t denote the cycle of length t. Then $e(G_n) = (\frac{1}{2}p + o(1))n^2$ and $C_t[G_n] = (p^t + o(1))n^t$.*

*Microsoft Research. Email: asafico@tau.ac.il.

[1] The result of [6] also contains quasi-random properties involving the number of induced copies of graphs of a fixed size, quasi-random properties related to the spectrum of the adjacency matrix of a graph, and more.

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

$\mathcal{P}_3$: *For any subset of vertices $U \subseteq V(G_n)$ of size $\frac{1}{2}n$ we have $e(U) = \frac{1}{2}p|U|^2 + o(n^2)$.*

$\mathcal{P}_4(\alpha)$: *Fix an $\alpha \in (0, \frac{1}{2})$. For any subset of vertices $U \subseteq V(G_n)$ of size αn we have that the number of edges connecting U and $V(G_n) \setminus U$ satisfies $e(U, V(G_n) \setminus U) = (p\alpha(1-\alpha) + o(1))n^2$.*

The meaning of the $o(1)$ terms in Theorem 1 is that for any $\delta > 0$ and large enough $n > n_0(\delta)$, G_n satisfies the appropriate condition up to δ. So a sequence of graphs (G_n) satisfies[2] property $\mathcal{P}_3$ if for any $\delta > 0$ and large enough $n > n_0(\delta)$, any subset of vertices $U \subseteq V(G_n)$ of size $\frac{1}{2}n$ satisfies $e(U) = \frac{1}{2}p|U|^2 \pm \delta n^2$. Also, the meaning of the fact that $\mathcal{P}_3$ implies $\mathcal{P}_2$ is that for any $\delta > 0$ there is an $\epsilon = \epsilon(\delta) > 0$ such that if G is an n-vertex graph with the property that any subset of vertices $U \subseteq V(G)$ of size $\frac{1}{2}n$ satisfies $e(U) = \frac{1}{2}p|U|^2 \pm \delta n^2$, then in fact any subset of vertices satisfies this condition. In what follows, let us say that a graph property is *quasi-random* if it is equivalent to any (and therefore all) of the 4 properties defined above. Note, that each of the four items in Theorem 1 is a property we would expect $G(n,p)$ to satisfy with high probability. Thus a quasi-random property asserts that G behaves like $G(n,p)$ (for the "right" p). Given Theorem 1 (and its omitted parts) one may stipulate that any property that holds for random graphs with high probability 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 $K_{n/2,n/2}$). As another example, note that in $\mathcal{P}_4$ we require $\alpha < \frac{1}{2}$, because when $\alpha = \frac{1}{2}$ the property is not quasi-random (see [6]). A more relevant family of non quasi-random properties are those requiring the graphs in the sequence to have the correct number of copies of a fixed graph H. Note that $\mathcal{P}_1(t)$ guarantees that for any even t, if a graph sequence has the correct number of edges as well as the correct number of copies of C_t then the sequence is quasi-random. As observed in [6] this is *not* true for all graphs.

Simonovits and Sós observed that the counter-examples showing that for some graphs H, having the correct number of copies of H 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 should be. As 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, they introduced the following variant of property $\mathcal{P}_1$ of Theorem 1, where now we require *all* subsets of vertices to contains the "correct" number of copies of H.

Definition 1.1 ($\mathcal{P}_H$) *For a fixed graph H on h vertices, we say that a graph sequence (G_n) satisfies $\mathcal{P}_H$ if all subsets of vertices $U \subseteq V(G_n)$ satisfy $H[U] = p^{e(H)}|U|^h + o(n^h)$.*

Note that the above condition does not impose any restriction on the number of edges of G, while in property $\mathcal{P}_1$ there is. As opposed to $\mathcal{P}_1$, which is quasi-random only for even-cycles, Simonovits and Sós [10] showed that $\mathcal{P}_H$ is quasi-random for any graph H.

Theorem 2 (Simonovits and Sós [10]) *For any fixed H with $e(H) > 0$, property $\mathcal{P}_H$ is quasi-random.*

[2]Here and throughout the paper we use the notation $x = y \pm z$ to denote that fact that $y - z \leq x \leq y + z$.

Note that property $\mathcal{P}_H$ is clearly not quasi-random if H is an edgeless graph, so one needs the assumption that $e(H) > 0$ in Theorem 2. We will henceforth assume that this is the case without explicitly mentioning this.

As we have argued above, the properties $\mathcal{P}_H$ can be thought of as a variant of property $\mathcal{P}_1$ in Theorem 1. But one can also think of the properties $\mathcal{P}_H$ as a generalization of property $\mathcal{P}_2$ in Theorem 1, namely, property $\mathcal{P}_2$ is just $\mathcal{P}_{K_{1,1}}$. Property $\mathcal{P}_3$ in Theorem 1 guarantees that in order to infer that a sequence is quasi-random, and thus satisfies $\mathcal{P}_2$, it is enough to require only the sets of vertices of size $n/2$ to contain the correct number of edges. At this point it is natural to ask if an analogous weaker condition also holds for the properties $\mathcal{P}_H$ when H is not an edge, that is, if it is enough to require that only sets of vertices of size n/c will have the correct number of copies of H. This question was first raised by Simonovits and Sós in [10]. Our main result here is a positive answer to this question. First let us define the following weaker variant of the properties $\mathcal{P}_H$:

Definition 1.2 ($\mathcal{P}'_H$) *For a fixed graph H on h vertices, we say that a graph sequence (G_n) satisfies $\mathcal{P}'_H$ if all subsets of vertices $U \subseteq V(G_n)$ of size $\lfloor n/(h+1) \rfloor$ satisfy $H[U] = p^{e(H)}|U|^h + o(n^h)$.*

Theorem 3 *For any fixed H, property $\mathcal{P}'_H$ is quasi-random.*

A natural open problem is if one can strengthen Theorem 3 by showing that it is in fact enough to consider sets of vertices of size n/c, for some absolute constant c that is independent of H. We conjecture that the answer is yes. In fact, it seems reasonable to conjecture that one can take c to be any constant larger than 1.

The main idea of the proof of Theorem 3 is to show that any graph that satisfies $\mathcal{P}'_H$ also satisfies $\mathcal{P}_H$. To this end we first show in Section 2 a simple equivalence between two types of properties regarding the number of copies of H in subsets of vertices. In Section 3 we apply an old algebraic result of Gottlieb [7] on the rank of Inclusion Matrices to show that certain sets of linear equations on hypergraphs are linearly independent. Finally, in Section 4 we apply the results of the first two sections in order to prove Theorem 3.

2 Copies of H vs Partite Copies of H

Recall that for a set of vertices $U \subseteq V$ we denote by $H[U]$ the number of labeled copies of H spanned by U. For an h-tuple of vertex sets $U_1, \ldots, U_h$ let us denote by $H[U_1, \ldots, U_h]$ the number of copies of H spanned by $U_1, \ldots, U_h$ that have precisely one vertex in each of the sets $U_1, \ldots, U_h$. We call such copies of H, partite copies of H with respect to $U_1, \ldots, U_h$. Our main goal in this section is to prove that the property of having the correct number of copies of a graph H in sets of vertices of size at least m is equivalent to the property of having the correct number of partite copies of H in h-tuples of sets of vertices of size at least m. This equivalence is proved in the following two lemmas.

Lemma 2.1 *For any $\delta > 0$ and any graph H on h vertices, there exists a $\gamma = \gamma_{2.1}(\delta, h) > 0$ such that the following holds: Suppose a graph G has the property that every set of vertices U of size $k \geq m$ satisfies $H[U] = k^h(p^{e(H)} \pm \gamma)$. Then, any h-tuple of sets of vertices $U_1, \ldots, U_h$ of size $k \geq m$ satisfies $H[U_1, \ldots, U_h] = h!k^h(p^{e(H)} \pm \delta)$.*

Proof: By the assumed properties of G we know that for any $1 \leq t \leq h$ and any subset $S \subseteq [h]$ of size t we have

$$H[\bigcup_{i \in S} U_i] = (tk)^h (p^{e(H)} \pm \gamma) \ . \tag{1}$$

By Inclusion-Exclusion we have that

$$\begin{aligned} H[U_1, \ldots, U_h] &= \sum_{t=h}^{1} (-1)^{h-t} \left(\sum_{S \subseteq [h] : |S| = t} H[\bigcup_{i \in S} U_i] \right) \\ &= \sum_{t=h}^{1} (-1)^{h-t} \binom{h}{t} (tk)^h (p^{e(H)} \pm \gamma) \\ &= k^h p^{e(H)} \sum_{t=h}^{1} \binom{h}{t} (-1)^{h-t} t^h \pm \gamma (2hk)^h \\ &= k^h h! p^{e(H)} \pm \gamma (2hk)^h \end{aligned}$$

where we have used (1) when moving from the first line to the second line, and the combinatorial identity $h! = \sum_{t=h}^{1} \binom{h}{t} (-1)^{h-t} t^h$ when moving from the third line to the fourth. Therefore, it is enough to take $\gamma = \gamma_{2.1}(\delta, h) = \delta/(2h)^h$. ■

Lemma 2.2 *For any $\delta > 0$ and any graph H on h vertices, there exists a $\gamma = \gamma_{2.2}(\delta) > 0$ such that the following holds for all large enough $k \geq k_{2.2}(\delta)$: Suppose a graph G has the property that every h-tuple of vertex sets $U_1, \ldots, U_h$ of size k/h are such that $H[U_1, \ldots, U_h] = h!(k/h)^h (p^{e(H)} \pm \gamma)$. Then every set of vertices U of size k is such that $H[U] = k^h (p^{e(H)} \pm \delta)$.*

Proof: Suppose G has a set of vertices U of size k for which $H[U] < k^h p^{e(H)} - \delta k^h$ (the case $H[U] > k^h p^{e(H)} + \delta k^h$ is identical). Define a partition of U into h sets $U_1, \ldots, U_h$ by taking a random permutation of the vertices of U, and defining U_1 as the first k/h vertices in the ordering, U_2 as the following k/h vertices, and so on. It is easy to see that for any fixed labelled copy of H in U the number of permutations for which the vertices of H belong to distinct sets U_i is precisely $h!(k-h)!(k/h)^h$. Hence, the probability that the vertices of H belong to distinct sets U_i is

$$h! \left(\frac{k}{h} \right)^h \frac{(k-h)!}{k!} = \frac{h!}{h^h} \cdot \frac{k^h}{k \cdot (k-1) \cdots (k-h+1)} \leq \frac{h!}{h^h} (1 + \frac{1}{2}\delta) \ ,$$

where the inequality is valid if $k \geq k_{2.2}(\delta)$ for an appropriate $k_{2.2}(\delta)$. As we assumed that U spans less than $k^h p^{e(H)} - \delta k^h$ labelled copies of H, we conclude by linearity of expectation that the expected number of labelled copies of H in such a partition is less than

$$\frac{h!}{h^h} (1 + \frac{1}{2}\delta) \cdot (k^h p^{e(H)} - \delta k^h) < h!(k/h)^h (p^{e(H)} - \gamma) \ ,$$

if we choose $\gamma = \gamma_{2.2}(\delta) = \frac{1}{2}\delta$. So there must be at least one permutation that defines a partition $U_1, \ldots, U_h$ with at most this many labelled copies of H, contradicting our assumption on G. ■

3 Inclusion Matrices, Linear Equations and Hypergraphs

Our main goal in this section is to prove Lemma 3.2 that will be used in the following section in the proof of Theorem 3. The proof of Lemma 3.2 will be a consequence of a result on the rank of inclusion matrices which we turn to define. As usual, let us denote the set of integers $\{1,\ldots,r\}$ by $[r]$. For integers $h \le d \le r$ let $I^r_{h,d}$ be the $\binom{r}{d} \times \binom{r}{h}$ matrix whose rows are indexed by d-element subsets of $[r]$ and whose columns are indexed by h-element subsets of $[r]$. With this convention, entry (S', S'') of $I^r_{h,d}$ is 1 if $S'' \subseteq S'$ and 0 otherwise. The following theorem of Gottlieb [7] (see also [1] and [13]) will be a central tool for us.

Theorem 4 (Gottlieb [7]) *For any $h \le d$ and $r \ge h + d$ the matrix $I^r_{h,d}$ has rank $\binom{r}{h}$.*

Our main goal in this section is to derive a *hypergraph* interpretation of the above result. We start with some basic definitions. An h-uniform hypergraph (h-graph for short) $\mathcal{H} = (V, E)$ has a set of vertices $V = V(\mathcal{H})$ and a set of edges $E = E(\mathcal{H})$ where every edge contains h distinct vertices from V. In what follows, we will identify V with the set of integers $[r]$ and denote by $\mathcal{H}_{r,h}$ the complete h-graph on r vertices.

Let $2 \le h \le d$ and $r \ge h + d$ be integers and let $\mathcal{D}$ be any h-graph on d vertices. Define $C^r_h(\mathcal{D})$ as the incidence matrix between the edges of $\mathcal{H}_{r,h}$ and the copies of $\mathcal{D}$ in $\mathcal{H}_{r,h}$. More precisely, let $c_{\mathcal{D}}$ denote the number of distinct copies of $\mathcal{D}$ that are spanned by d vertices of $\mathcal{H}_{r,h}$ (that is, by $\mathcal{H}_{d,h}$), and let $C^r_h(\mathcal{D})$ be a $c_{\mathcal{D}} \cdot \binom{r}{d} \times \binom{r}{h}$ matrix whose $c_{\mathcal{D}} \cdot \binom{r}{d}$ rows are indexed by the copies of $\mathcal{D}$ in $\mathcal{H}_{r,h}$, and whose $\binom{r}{h}$ columns are indexed by edges of $\mathcal{H}_{r,h}$. Then, the entry of $C^r_h(\mathcal{D})$ corresponding to a copy of $\mathcal{D}$ and an edge $e \in E(\mathcal{H}_{r,h})$ is 1 if $e \in E(\mathcal{D})$ and is 0 otherwise.

Lemma 3.1 *For any triple of integers $2 \le h \le d$ and $r \ge h + d$, and any h-graph $\mathcal{D}$ on d vertices, the matrix $C^r_h(\mathcal{D})$ has rank $\binom{r}{h}$.*

Proof: Observe that when $\mathcal{D} = \mathcal{H}_{d,h}$ (i.e. the complete h-graph on d vertices) then $c_{\mathcal{D}} = 1$ and therefore $C^r_h(\mathcal{H}_{d,h})$ is an $\binom{r}{d} \times \binom{r}{h}$ matrix. Furthermore, if we identify a copy of $\mathcal{H}_{d,h}$ with its set of vertices then one can easily see that $C^r_h(\mathcal{H}_{d,h})$ is in fact identical to the matrix $I^r_{h,d}$. Theorem 4 thus implies that $C^r_h(\mathcal{H}_{d,h})$ has rank $\binom{r}{h}$. Consider now an arbitrary h-graph $\mathcal{D}$ on d vertices. Let v_S be the row of $C^r_h(\mathcal{H}_{d,h}))$ corresponding to the copy of $\mathcal{H}_{d,h}$ on the set of vertices S. For each $S \subseteq [r]$ of size d, let u_S be the row vector obtained by taking the sum of the $c_{\mathcal{D}}$ rows of $C^r_h(\mathcal{D})$ corresponding to the copies of $\mathcal{D}$ spanned by S. By symmetry, every edge of $\mathcal{H}_{r,h}$ in S is contained in the same number of copies of $\mathcal{D}$ that are spanned by S. Therefore, $u_S = c'_{\mathcal{D}} \cdot v_S$ for some constant $c'_{\mathcal{D}}$. Hence, the rows of $C^r_h(\mathcal{D})$ span the rows of $C^r_h(\mathcal{H}_{d,h})$, so the rank of $C^r_h(\mathcal{D})$ must be $\binom{r}{h}$. ■

For our purposes we will in fact be interested in one type of h-graph $\mathcal{D}$. Let $\mathcal{D}_{h,b}$ be the complete h-partite h-graph on bh vertices whose every partition class contains b vertices. In other words, we think of the vertex set of $\mathcal{D}_{h,b}$ as being composed of h sets of vertices $S_1, \ldots, S_h$ of size b each. The edge set of $\mathcal{D}_{h,b}$ has all sets of h vertices that contain one vertex from each of the sets S_i (so $\mathcal{D}_{h,b}$ has b^h edges). It will be convenient to refer to a copy of $\mathcal{D}_{h,b}$ by referring to its h vertex sets $S_1, \ldots, S_h$. Note that by Lemma 3.1 we have that for any $h \ge 2$, $b \ge 1$ the matrix $C^{bh+h}_h(\mathcal{D}_{h,b})$ has rank $\binom{r}{h}$. The following is the main result of this section.

Lemma 3.2 *Let $\mathcal{H} = \mathcal{H}_{bh+h,h}$ be the complete h-graph on $bh+h$ vertices, let $\mathcal{D} = \mathcal{D}_{h,b}$ be the complete h-partite h-graph on bh vertices. Assign to each edge $e = \{v_1, \ldots, v_h\}$ of $\mathcal{H}$ an unknown variable $x_{v_1,\ldots,v_h}$. Then $\mathcal{H}$ contains $t = \binom{bh+h}{h}$ copies of $\mathcal{D}$, denoted $\mathcal{D}_1, \ldots, \mathcal{D}_t$, with the following property: suppose that for every copy $\mathcal{D}_j$, which is spanned by $S_1^j, \ldots, S_h^j$, we write a linear equation*

$$\sum_{v_1 \in S_1^j, \ldots, v_h \in S_h^j} x_{v_1,\ldots,v_h} = b_j \ , \tag{2}$$

where b_j is an arbitrary real. Then this system of linear equations has a unique solution.

Proof: This system of linear equations has $\binom{bh+h}{h}$ unknowns (one per edge of $\mathcal{H}$) and $t = \binom{bh+h}{h}$ equations (one per copy $\mathcal{D}_j$). If we write it as $Ax = b$, then we only need to show that we can choose the copies of $\mathcal{D}$ in such a way that A is non singular. The important observation now is that the row of $C_h^{bh+h}(\mathcal{D}_{h,b})$ corresponding to a copy of $\mathcal{D}$ is the same as the row of A corresponding to the same copy of $\mathcal{D}$. This is because for a copy $\mathcal{D}_j$ of $\mathcal{D}$, the linear equation in (2) involves the unknowns $x_{i_1,\ldots,i_h}$ that correspond to the edges $\{i_1, \ldots, i_h\}$ of $\mathcal{D}_j$. By Lemma 3.1 we know that $C_h^{bh+h}(\mathcal{D}_{h,b})$ has rank $\binom{bh+h}{h}$, hence, it has a set of $\binom{bh+h}{h}$ rows that are linearly independent. Therefore, if we use the $\binom{bh+h}{h}$ copies of $\mathcal{D}$ that correspond to these rows we get that the matrix A is non-singular. ■

4 Proof of Main Result

In this section we combine the results from the previous sections and prove Theorem 3. One additional ingredient we will need, is the following simple claim.

Claim 4.1 *For any integer p there is a $C = C_{4.1}(p)$ 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 $\ell_\infty(b', b) \le \gamma$ then the unique solution x' of $Ax' = b'$ satisfies $\ell_\infty(x', x) \le C\gamma$.*

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 linear function of b it is clear that there is a $C = C(A)$ such that if $\ell_\infty(b', b) \le \gamma$ then the unique solution x' of $Ax' = b'$ satisfies $\ell_\infty(x', x) \le C\gamma$. Now, as there are finitely many $0/1$ $p \times p$ matrices, we can set $C = C_{4.1}(p) = \max_A C(A)$, where the maximum is taken over all $0/1$ $p \times p$ matrices. ■

Proof of Theorem 3: Fix any graph H on h vertices. Our goal is to show that any graph sequence that satisfies $\mathcal{P}'_H$ also satisfies $\mathcal{P}_H$ and is thus quasi-random by Theorem 2. Recall that $\mathcal{P}_H$ is equivalent to requiring that for all $\delta > 0$ and for all large enough $n > n_0(\delta)$, all subsets of vertices $U \subseteq V(G_n)$ must satisfy $H[U] = p^{e(H)}|U|^h \pm \delta n^h$. As this condition vacuously holds for sets of size at most δn, it is enough to show that for all $\delta > 0$ and large enough $n > n_0(\delta)$, all subsets of vertices $U \subseteq V(G_n)$ of size at leats δn satisfy the stronger condition

$$H[U] = (p^{e(H)} \pm \delta)|U|^h \ . \tag{3}$$

If fact, by averaging, if (3) holds for sets of size δn then it holds also for all sets of size at least δn (with a slightly larger δ). Fix then any $\delta > 0$, and assume wlog that $\delta < \frac{1}{h+1}$. Define the following constants $\gamma_{2.2} = \gamma_{2.2}(\delta)$, $\gamma_{4.1} = \gamma_{2.2}/((1/\delta - 1)^h \cdot C_{4.1}(\binom{h/\delta}{h}))$ and $\gamma_3 = \gamma_{2.1}(\gamma_{4.1}, h)$. As (G_n) is assumed to satisfy $\mathcal{P}'_H$, we know that for any $\gamma > 0$, any large enough graph in the sequence, has the property that all subsets of vertices U of size $|V(G_n)|/(h+1)$ satisfy $H[U] = p^{e(H)}|U|^h \pm \gamma n^h$. This is equivalent to saying that (for a slightly smaller γ) such sets satisfy

$$H[U] = (p^{e(H)} \pm \gamma)|U|^h \ . \tag{4}$$

We claim that any graph $G \in (G_n)$ that satisfies (4) with γ_3 (defined above) also satisfies (3) and thus by the above discussion (G_n) satisfies $\mathcal{P}_H$. By Lemma 2.2 we know that in order to derive that any subset U of size δn satisfies (3) it is enough to prove that every h-tuple of vertex sets $U_1, \ldots, U_h$ of size $\delta n/h$ satisfies $H[U_1, \ldots, U_h] = h!(\delta n/h)^h(p^{e(H)} \pm \gamma_{2.2})$. So consider any h-tuples of sets $U_1, \ldots, U_h$ of size $\delta n/h$ each and let us partition the rest of the vertices of G into sets $U_{h+1}, \ldots, U_{h/\delta}$ of size $\delta n/h$ each (we assume wlog that h/δ is an integer). For the rest of the proof, let us denote the size of the sets U_i by $m = \frac{\delta n}{h}$. We will show that in fact *all* h-tuples $U_{i_1}, \ldots, U_{i_h}$ in this partition have $h!m^h(p^{e(H)} \pm \gamma_{2.2})$ partite copies of H, that is, that

$$H(U_{i_1}, \ldots, U_{i_h}) = h!m^h(p^{e(H)} \pm \gamma_{2.2}) \ . \tag{5}$$

Our assumption is that every set of vertices $U \subseteq V(G)$ of size $n/(h+1)$ satisfies (4) with $\gamma = \gamma_3$. Hence [3], we have through Lemma 2.1 that every h tuple of vertices, each of size $k \geq n/(h+1)$, spans $h!k^h(p^{e(H)} \pm \gamma_{4.1})$ copies of H. We now turn to use this fact to derive (5).

Recall that we have a partition of G into h/δ sets U_i each of size $m = \delta n/h$. For convenience, let us define b as an integer for which $bh + h = h/\delta$, that is $b = \frac{1}{\delta} - 1$. Define a *super partition* of the $bh + h$ sets U_i as a partition that clusters bh of the sets into h supersets $S_1, \ldots, S_h$ of size b each (the other h sets are ignored in this super partition). Given a super partition $S_1, \ldots, S_h$ of the $bh + b$ sets U_i, let us define the corresponding super partition of $V(G)$ into h sets $V_1, \ldots, V_h$ by setting $V_i = \bigcup_{t \in S_i} U_t$. As each of the supersets S_i contains b of the sets U_i, we get that each of the sets V_i is of size $bm \geq n/(h+1)$ (here we use $\delta < \frac{1}{h+1}$). Thus, the properties of G guarantee that for any super partition $V_1, \ldots, V_h$

$$H(V_1, \ldots, V_h) = h!(bm)^h(p^{e(H)} \pm \gamma_{4.1}) \ . \tag{6}$$

We can also count the number of copies of H with one vertex in each set V_i by considering the number of copies of H with one vertex in one of the sets $U_t \subseteq V_i$. That is, we can rewrite (6) as

$$H(V_1, \ldots, V_h) = \sum_{i_1 \in S_1, \ldots, i_h \in S_h} H(U_{i_1}, \ldots, U_{i_h}) = h!(bm)^h(p^{e(H)} \pm \gamma_{4.1}) \ . \tag{7}$$

Let us set

$$x_{i_1,\ldots,i_h} = \frac{1}{(bm)^h h!} \cdot H(U_{i_1}, \ldots, U_{i_h}) \ . \tag{8}$$

[3]Here again we use the simple observation that if all the sets U of size $n/(h+1)$ satisfy (4) then by averaging the same applies for *larger* sets.

Then, for any super partition $S_1, \ldots, S_h$ we can rewrite (7) as follows

$$\sum_{i_1 \in S_1, \ldots, i_h \in S_h} x_{i_1,\ldots,i_h} = \frac{1}{(bm)^h h!} \cdot H(V_1, \ldots, V_h) = p^{e(H)} \pm \gamma_{4.1} \; . \tag{9}$$

Recall that we have a partition of G into $bh+b$ sets U_i, a super partition of these $bh+b$ sets into h clusters $S_1, \ldots, S_h$ and that we have assigned a number $x_{i_1,\ldots,i_h}$ to each subset of h of these clusters. Let $\mathcal{H}$ be the complete h-graph on $bh+b$ vertices, and think of each $x_{i_1,\ldots,i_h}$ as an unknown assigned to the hyper edge containing the vertices $\{i_1, \ldots, i_h\}$. The important observation at this point is that each of the linear equations in (9) is exactly the linear equation in (2) when taking the copy of $\mathcal{D}$ in $\mathcal{H}$ to be the one spanned by $S_1, \ldots, S_h$, and b_j to be $\frac{1}{(bm)^h h!} \cdot H(V_1, \ldots, V_h) = p^{e(H)} \pm \gamma_{4.1}$. Lemma 3.2 thus guarantees that there are $\binom{bh+h}{h}$ super partitions of the $bh+h$ sets U_i such that the corresponding $\binom{bh+h}{h}$ linear equations as in (9) have a unique solution. Note that each of the equations in (9) involves b^h terms $x_{i_1,\ldots,i_h}$. Therefore, if we had $\gamma_{4.1} = 0$ in all the RHSs of the linear equations, then the unique solution would have been $x_{i_1,\ldots,i_h} = p^{e(H)}/b^h$ for all $\{i_1, \ldots, i_h\}$. Now, as the system involves $\binom{bh+h}{h} = \binom{h/\delta}{h}$ linear equations, and $\gamma_{4.1} = \gamma_{2.2}/(b^h \cdot C_{4.1}(\binom{h/\delta}{h}))$ we have through Claim 4.1 that the unique solution to the system of linear equations satisfies $x_{i_1,\ldots,i_h} = p^{e(H)}/b^h \pm \gamma_{2.2}/b^h$ for all $\{i_1, \ldots, i_h\}$. Recalling (8), this means that for all $i_1, \ldots, i_h$, the number of partite copies of H spanned by $U_{i_1}, \ldots, U_{i_h}$ satisfies

$$H(U_{i_1}, \ldots, U_{i_h}) = h!(bm)^h \cdot x_{i_1,\ldots,i_h} = h! m^h (p^{e(H)} \pm \gamma_{2.2}) \; ,$$

thus completing the proof. ■

Acknowledgements: I would to thank Miki Simonovits, Vera T. Sós and Asaf Nachmias for helpful discussions. Special thanks to Raphy Yuster for a careful reading of this paper and for many helpful comments.

References

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

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

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

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

[5] 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.

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

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

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

[9] L. Lovász and V. T. Sós, Generalized quasirandom graphs, Journal of Combinatorial Theory Series B 98 (2008), 146-163.

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

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

[12] 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.

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