Quasi-randomness is determined by the distribution of copies of a fixed graph in equicardinal large sets

Raphael Yuster *

Abstract

For every fixed graph $H$ and every fixed $0 < \alpha < 1$, we show that if a graph $G$ has the property that all subsets of size $\alpha n$ contain the “correct” number of copies of $H$ one would expect to find in the random graph $G(n, p)$ then $G$ behaves like the random graph $G(n, p)$; that is, it is $p$-quasi-random in the sense of Chung, Graham, and Wilson [4]. This solves a conjecture raised by Shapira [8] and solves in a strong sense an open problem of Simonovits and Sós [9].

1 Introduction

The theory of quasi-random graphs asks the following fundamental question: which properties of graphs are such that any graph that satisfies them, resembles an appropriate random graph (namely, the graph satisfies the properties that a random graph would satisfy, with high probability). Such properties are called quasi-random.

The theory of quasi-random graphs was initiated by Thomason [10, 11] and then followed by Chung, Graham and Wilson who proved the fundamental theorem of quasi-random graphs [4]. Since then there have been many papers on this subject (see, e.g. the excellent survey [6]). Quasi-random properties were also studied for other combinatorial structures such as set systems [1], tournaments [2], and hypergraphs [3]. There are also some very recent results on quasi-random groups [5] and generalized quasi-random graphs [7].

In order to formally define $p$-quasi-randomness we need to state the fundamental theorem of quasi-random graphs. As usual, a labeled copy of a graph $H$ in a graph $G$ is an injective mapping $\phi$ from $V(H)$ to $V(G)$ that maps edges to edges. That is $(x, y) \in E(H)$ implies $(\phi(x), \phi(y)) \in E(G)$. For a set of vertices $U \subset V(G)$ we denote by $H[U]$ the number of labeled copies of $H$ in the subgraph of $G$ induced by $U$ and by $e(U)$ the number of edges of $G$ with both endpoints in $U$. A graph sequence $(G_n)$ is an infinite sequence of graphs $\{G_1, G_2, \ldots\}$ where $G_n$ has $n$ vertices. The following result of Chung, Graham, and Wilson [4] shows that many properties of different nature

*Department of Mathematics, University of Haifa, Haifa 31905, Israel. E-mail: raphy@math.haifa.ac.il
are equivalent to the notion of quasi-randomness, defined using edge distribution. The original theorem lists seven such equivalent properties, but we only state four of them here.

**Theorem 1 (Chung, Graham, and Wilson [4])** Fix any $1 < p < 1$. For any graph sequence $(G_n)$ the following properties are equivalent:

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

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

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

- **P$_4(\alpha)$:** Fix an $\alpha \in (0, \frac{1}{2})$. For any $U \subseteq V(G_n)$ of size $\alpha n$ we have $e(U, V \setminus U) = p\alpha(1 - \alpha)n^2 + o(n^2)$.

The formal meaning of the properties being equivalent is expressed, as usual, using $\epsilon, \delta$ notation. For example the meaning that $P_3$ implies $P_2$ is that for any $\epsilon > 0$ there exist $\delta = \delta(\epsilon)$ and $N = N(\epsilon)$ so that for all $n > N$, if $G$ is a graph with $n$ vertices having the property that any subset of vertices $U$ of size $n/2$ satisfies $|e(U) - \frac{1}{2}p|U|^2| < \delta n^2$ then also for any subset of vertices $W$ we have $|e(W) - \frac{1}{2}p|W|^2| < \epsilon n^2$.

Given Theorem 1 we say that a graph property is $p$-quasi-random if it is equivalent to any (and therefore all) of the four properties defined in that theorem. (We will usually just say quasi-random instead of $p$-quasi-random since $p$ is fixed throughout the proofs). Note, that each of the four properties in Theorem 1 is a property we would expect $G(n, p)$ to satisfy with high probability.

It is far from true, however, that any property that almost surely holds for $G(n, p)$ is quasi-random. For example, it is easy to see that having vertex degrees $np(1 + o(1))$ is not a quasi-random property (just take vertex-disjoint cliques of size roughly $np$ each). An important 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 $P_1(t)$ 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 $H = C_t$ then the sequence is quasi-random. As observed in [4] this is not true for all graphs. In fact, already for $H = K_3$ there are simple constructions showing that this is not true.

Simonovits and Sós observed that the standard counter-examples showing that for some graphs $H$, having the correct number of copies of $H$ is not enough to guarantee quasi-randomness, have the property that the number of copies of $H$ in some of the induced subgraphs of these counter-examples deviates significantly from what it 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 $P_1$ 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 $H$ with $h$ vertices and $r$ edges, we say that a graph sequence $(G_n)$ satisfies $\mathcal{P}_H$ if all subsets of vertices $U \subset V(G_n)$ with $|U| = \alpha n$ satisfy $H[U] = p|U|^h + o(n^h)$.

As opposed to $\mathcal{P}_1$, which is quasi-random only for even cycles, Simonovits and Sós [9] showed that $\mathcal{P}_H$ is quasi-random for any nonempty graph $H$.

Theorem 2 For any fixed $H$ that has edges, property $\mathcal{P}_H$ is quasi-random.

We can view property $\mathcal{P}_H$ as a generalization of property $\mathcal{P}_2$ in Theorem 1 since $\mathcal{P}_2$ is just the special case $\mathcal{P}_{K_2}$. Now, 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. An open problem raised by Simonovits and Sós [9], and in a stronger form by Shapira [8], is that the analogous condition also holds for any $H$. Namely, in order to infer that a sequence is quasi-random, and thus satisfies $\mathcal{P}_H$, it is enough, say, to require only the sets of vertices of size $n/2$ to contain the correct number of copies of $H$. Shapira [8] proved that it is enough to consider sets of vertices of size $n/(h+1)$. Hence, in his result, the cardinality of the sets depends on $h$. Thus, if $H$ has 1000 vertices, Shpiras’s result shows that it suffices to check vertex subsets having a fraction smaller than $1/1000$ of the total number of vertices. His proof method cannot be extended to obtain the same result for fractions larger than $1/(h + \epsilon)$.

In this paper we settle the above mentioned open problem completely. In fact, we show that for any $H$, not only is it enough to check only subsets of size $n/2$, but, more generally, we show that it is enough to check subsets of size $\alpha n$ for any fixed $\alpha \in (0, 1)$. More formally, we define:

Definition 1.2 ($\mathcal{P}_{H,\alpha}$) For a fixed graph $H$ with $h$ vertices and $r$ edges and fixed $0 < \alpha < 1$ we say that a graph sequence $(G_n)$ satisfies $\mathcal{P}_{H,\alpha}$ if all subsets of vertices $U \subset V(G_n)$ with $|U| = \lfloor \alpha n \rfloor$ satisfy $H[U] = p'|U|^h + o(n^h)$.

Our main result is, therefore:

Theorem 3 For any fixed graph $H$ and any fixed $0 < \alpha < 1$, property $\mathcal{P}_{H,\alpha}$ is quasi-random.

2 Proof of the main result

For the remainder of this section let $H$ be a fixed graph with $h > 1$ vertices and $r > 0$ edges, and let $\alpha \in (0, 1)$ be fixed. Throughout this section we ignore rounding issues and, in particular, assume that $\alpha n$ is an integer, as this has no effect on the asymptotic nature of the results.

Suppose that the graph sequence $(G_n)$ satisfies $\mathcal{P}_{H,\alpha}$. We will prove that it is quasi-random by showing that it also satisfies $\mathcal{P}_H$. In other words, we need to prove the following lemma which, together with Theorem 2 yields Theorem 3.
**Lemma 2.1** For any $\epsilon > 0$ there exists $N = N(\epsilon, h, \alpha)$ and $\delta = \delta(\epsilon, h, \alpha)$ so that for all $n > N$, if $G$ is a graph with $n$ vertices satisfying that for all $U \subseteq V(G)$ with $|U| = \alpha n$ we have $|H[U] - p^r|U|^h| < \delta n^h$ then $G$ also satisfies that for all $W \subseteq V(G)$ we have $|H[W] - p^r|W|^h| < \epsilon n^h$.

**Proof:** Suppose therefore that $\epsilon > 0$ is given. Let $N = N(\epsilon, h, \alpha)$, $\epsilon' = \epsilon'(\epsilon, h, \alpha)$ and $\delta = \delta(\epsilon, h, \alpha)$ be parameters to be chosen so that $N$ is sufficiently large and $\delta \ll \epsilon'$ are both sufficiently small to satisfy the inequalities that will follow, and it will be clear that they are indeed only functions of $\epsilon, h$, and $\alpha$.

Now, let $G$ be a graph with $n > N$ vertices satisfying that for all $U \subseteq V(G)$ with $|U| = \alpha n$ we have $|H[U] - p^r|U|^h| < \delta n^h$. Consider any subset $W \subseteq V(G)$. We need to prove that $|H[W] - p^r|W|^h| < \epsilon n^h$.

For convenience, set $k = \alpha n$. Let us first prove this for the case where $|W| = m > k$. This case can rather easily be proved via a simple counting argument. Denote by $\mathcal{U}$ the set of $\binom{m}{k}$ $k$-subsets of $W$. Hence, by the given condition on $k$-subsets,

$$
\binom{m}{k}(p^r k^h - \delta n^h) < \sum_{U \in \mathcal{U}} H[U] < \binom{m}{k}(p^r k^h + \delta n^h). \tag{1}
$$

Every copy of $H$ in $W$ appears in precisely $\binom{m-h}{k-h}$ distinct $U \in \mathcal{U}$. it follows from (1) that

$$
H[W] = \frac{1}{\binom{m-h}{k-h}} \sum_{U \in \mathcal{U}} H[U] < \frac{\binom{m}{k}}{\binom{m-h}{k-h}}(p^r k^h + \delta n^h) < p^r m^h + \frac{\epsilon'}{2} n^h, \tag{2}
$$

and similarly from (1)

$$
H[W] = \frac{1}{\binom{m-h}{k-h}} \sum_{U \in \mathcal{U}} H[U] > \frac{\binom{m}{k}}{\binom{m-h}{k-h}}(p^r k^h - \delta n^h) > p^r m^h - \frac{\epsilon'}{2} n^h. \tag{3}
$$

We now consider the case where $|W| = m = \beta n < \alpha n = k$. Notice that we can assume that $\beta \geq \epsilon$ since otherwise the result is trivially true. The set $\mathcal{H}$ of $H$-subgraphs of $G$ can be partitioned into $h + 1$ types, according to the number of vertices they have in $W$. Hence, for $j = 0, \ldots, h$ let $\mathcal{H}_j$ be the set of $H$-subgraphs of $G$ that contain precisely $j$ vertices in $V \setminus W$. Notice that, by definition, $|\mathcal{H}_0| = H[W]$. For convenience, denote $w_j = |\mathcal{H}_j|/n^h$. We therefore have

$$
w_0 + w_1 + \cdots + w_h = \frac{|\mathcal{H}|}{n^h} = \frac{H[V]}{n^h} = p^r + \mu \tag{4}
$$

where $|\mu| < \epsilon'/2$.

Define $\lambda = \frac{(1-\alpha)}{h+1}$ and set $k_i = k + i\lambda n$ for $i = 1, \ldots, h$. Let $Y_i \subseteq V \setminus W$ be a random set of $k_i - m$ vertices, chosen uniformly at random from all $\binom{n-m}{k_i-m}$ subsets of size $k_i - m$ of $V \setminus W$. Denote $K_i = Y_i \cup W$ and notice that $|K_i| = k_i > \alpha n$. We will now estimate the number of elements of $\mathcal{H}_j$.
that “survive” in $K_i$. Formally, let $\mathcal{H}_{j,i}$ be the set of elements of $\mathcal{H}_j$ that have all of their vertices in $K_i$, and let $m_{j,i} = |\mathcal{H}_{j,i}|$. Clearly, $m_{0,i} = H[W]$ since $W \subset K_i$. Furthermore, by (2) and (3),

\[ m_{0,i} + m_{1,i} + \cdots + m_{h,i} = H[K_i] = p^r k_i^h + \rho_i n^h \tag{5} \]

where $\rho_i$ is a random variable with $|\rho_i| < \epsilon'/2$.

For an $H$-copy $T \in \mathcal{H}_j$ we compute the probability $p_{j,i}$ that $T \in H[K_i]$. Since $T \in H[K_i]$ if and only if all the $j$ vertices of $T$ in $V \setminus W$ appear in $Y_i$ we have

\[ p_{j,i} = \frac{\binom{n-m-j}{k_i-m-j}}{\binom{n-m}{k_i-m}} = \frac{(k_i - m) \cdots (k_i - m - j + 1)}{(n-m) \cdots (n-m-j+1)}. \]

Defining $x_i = (k_i - m)/(n - m)$ and noticing that

\[ x_i = \frac{k_i - m}{n - m} = \frac{\alpha - \beta}{1 - \beta} + \frac{\lambda}{1 - \beta} i \]

it follows that

\[ |p_{j,i} - x_i|^j < \frac{\epsilon'}{2}. \tag{6} \]

Clearly, the expectation of $m_{j,i}$ is $E[m_{j,i}] = p_{j,i} |\mathcal{H}_j|$. By linearity of expectation we have from (5) that

\[ E[m_{0,i}] + E[m_{1,i}] + \cdots + E[m_{h,i}] = E[H[K_i]] = p^r k_i^h + E[\rho_i] n^h. \]

Dividing the last equality by $n^h$ we obtain

\[ p_{0,i} w_0 + \cdots + p_{h,i} w_h = p^r (\alpha + \lambda i)^h + E[\rho_i]. \tag{7} \]

By (6) and (7) we therefore have

\[ \sum_{j=0}^h x_i^j w_j = p^r (\alpha + \lambda i)^h + \mu_i \tag{8} \]

where $\mu_i = E[\rho_i] + \zeta_i$ and $|\zeta_i| < \epsilon'/2$. Since also $|\rho_i| < \epsilon'/2$ we have that $|\mu_i| < \epsilon'$.

Now, (4) and (8) form together a system of $h + 1$ linear equations with the $h + 1$ variables $w_0, \ldots, w_h$. The coefficient matrix of this system is just the Vandermonde matrix $A = A(x_1, \ldots, x_h, 1)$. Since $x_1, \ldots, x_h, 1$ are all distinct, and, in fact, the gap between any two of them is at least $\lambda/(1 - \beta) = (1 - \alpha)/(h + 1)(1 - \beta)) \geq (1 - \alpha)/(h + 1)$, we have that the system has a unique solution which is $A^{-1} b$ where $b \in R^{h+1}$ is the column vector whose $i$'th coordinate is $p^r (\alpha + \lambda i)^h + \mu_i$ for $i = 1, \ldots, h$ and whose last coordinate is $p^r + \mu$. Consider now the vector $b^*$ which is the same as $b$, just without the $\mu_i$’s. Namely $b^* \in R^{h+1}$ is the column vector whose $i$'th coordinate is $p^r (\alpha + \lambda i)^h$ for $i = 1, \ldots, h$ and whose last coordinate is $p^r$. Then the system $A^{-1} b^*$
also has a unique solution and, in fact, we know exactly what this solution is. It is the vector 
\( w^* = (w_0^*, \ldots, w_h^*) \) where 
\[
w_j^* = p^r \binom{h}{j} \beta^{h-j}(1-\beta)^j.
\]
Indeed, it is straightforward to verify the equality
\[
\sum_{j=0}^{h} p^r \binom{h}{j} \beta^{h-j}(1-\beta)^j = p^r
\]
and, for all \( i = 1, \ldots, h \) the equalities
\[
\sum_{j=0}^{h} \left( \frac{\alpha - \beta}{1-\beta} + \frac{\lambda}{1-\beta} \right)^j p^r \binom{h}{j} \beta^{h-j}(1-\beta)^j = p^r (\alpha + \lambda h)^h.
\]
Now, since the mapping \( F : R^{h+1} \to R^{h+1} \) mapping a vector \( c \) to \( A^{-1}c \) is continuous, we know that for \( \epsilon' \) sufficiently small, if each coordinate of \( c \) has absolute value less than \( \epsilon' \), then each coordinate of \( A^{-1}c \) has absolute value at most \( \epsilon \). Now, define \( c = b - b^* = (\mu_1, \ldots, \mu_h, \mu) \). Then we have that each coordinate \( w_i \) of \( A^{-1}b \) differs from the corresponding coordinate \( w_i^* \) of \( A^{-1}b^* \) by at most \( \epsilon \). In particular,
\[
|w_0 - w_0^*| = |w_0 - p^r \beta^h| < \epsilon.
\]
Hence,
\[
|H[W] - n^h p^r \beta^h| = |H[W] - p^r |W|^h| < \epsilon n^h
\]
as required.

References

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

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

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

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

[5] T. Gowers. Quasirandom groups. manuscript, 2006.
[6] M. Krivelevich and B. Sudakov. Pseudo-random graphs. *More sets, graphs and numbers*, E. Györi, G.O.H. Katona and L. Lovász, Eds., Bolyai Society Mathematical Studies Vol. 15, 199–262.

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

[8] A. Shapira. Quasi-randomness and the distribution of copies of a fixed graph. *Combinatorica*, to appear.

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

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

[11] 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.