Quasi-Randomness of Graph Balanced Cut Properties

Hao Huang, Choongbum Lee
Department of Mathematics, UCLA, Los Angeles, California 90095; e-mail: huanghao@math.ucla.edu; abdesire@math.ucla.edu

Received 13 September 2010; accepted 6 May 2011
Published online 29 September 2011 in Wiley Online Library (wileyonlinelibrary.com).
DOI 10.1002/rsa.20384

ABSTRACT: Quasi-random graphs can be informally described as graphs whose edge distribution closely resembles that of a truly random graph of the same edge density. Recently, Shapira and Yuster proved the following result on quasi-randomness of graphs. Let $k \geq 2$ be a fixed integer, $\alpha_1, \ldots, \alpha_k$ be positive reals satisfying $\sum \alpha_i = 1$ and $(\alpha_1, \ldots, \alpha_k) \neq (1/k, \ldots, 1/k)$, and $G$ be a graph on $n$ vertices. If for every partition of the vertices of $G$ into sets $V_1, \ldots, V_k$ of size $\alpha_1n, \ldots, \alpha_kn$, the number of complete graphs on $k$ vertices which have exactly one vertex in each of these sets is similar to what we would expect in a random graph, then the graph is quasi-random. However, the method of quasi-random hypergraphs they used did not provide enough information to resolve the case $(1/k, \ldots, 1/k)$ for graphs. In their work, Shapira and Yuster asked whether this case also forces the graph to be quasi-random. Janson also posed the same question in his study of quasi-randomness under the framework of graph limits. In this paper, we positively answer their question. © 2011 Wiley Periodicals, Inc. Random Struct. Alg., 41, 124–145, 2012

Keywords: quasi-random graphs; pseudo-random graphs; cut properties

1. INTRODUCTION

The study of random structures has seen a tremendous success in modern combinatorics and theoretical computer science. One example is the Erdős-Rényi random graph $G(n, p)$ proposed in the 1950’s and intensively studied thereafter. $G(n, p)$ is the probability space of graphs over $n$ vertices where each pair of vertices forms an edge independently with probability $p$. Random graphs are not only an interesting object of study on their own but also proved to be a powerful tool in solving numerous open problems. The success of random structures served as a natural motivation for the following question: How can one tell when...
a given structure behaves like a random one? Such structures are called quasi-random. In this paper we study quasi-random graphs, which, following Thomason \[18, 19\], can be informally defined as graphs whose edge distribution closely resembles that of a random graph (the formal definition will be given later). One fundamental result in the study of quasi-random graphs is the following theorem proved by Chung, Graham and Wilson \[3\] (here we only state part of their result).

**Theorem 1.1.** Fix a real \( p \in (0, 1) \). For an \( n \)-vertex graph \( G \), define \( e(U) \) to be the number of edges in the induced subgraph spanned by vertex set \( U \), then the following properties are equivalent.

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

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

\( \mathcal{P}_3 \): \( e(G) = \frac{1}{2}pn^2 + o(n^2) \) and \( G \) has \( \frac{1}{2}p^4n^4 \pm o(n^4) \) cycles of length 4.

Throughout this paper, unless specified otherwise, when considering a subset of vertices \( U \subseteq V \) such that \( |U| = \alpha n \) for some \( \alpha \), we tacitly assume that \( |U| = \lfloor \alpha n \rfloor \) or \( |U| = \lceil \alpha n \rceil \). Since we mostly consider asymptotic values, this difference will not affect our calculation.

For a positive real \( \delta \), we say that a graph \( G \) is \( \delta \)-close to satisfying \( \mathcal{P}_1 \) if \( e(U) = \frac{1}{2}p|U|^2 \pm \delta n^2 \) for all \( U \subseteq V(G) \), and similarly define it for other properties. The formal definition of equivalence of properties in Theorem 1.1 is as following: for every \( \epsilon > 0 \), there exists a \( \delta \) such that if a graph is \( \epsilon \)-close to satisfying one property, then it is \( \epsilon \)-close to satisfying another.

We call a graph \( p \)-quasi-random, or quasi-random if the density \( p \) is clear from the context, if it satisfies \( \mathcal{P}_1 \), and consequently satisfies all of the equivalent properties of Theorem 1.1. We also say that a graph property is quasi-random if it is equivalent to \( \mathcal{P}_1 \). Note that the random graph \( G(n, p) \) with high probability is \( p \)-quasi-random. However, it is not true that all the properties of random graphs are quasi-random. For example, it is easy to check that the property of having \( \frac{1}{2}p(n^2) + o(n^2) \) edges is not quasi-random (as an instance, there can be many isolated vertices). For more details on quasi-random graphs we refer the reader to the survey of Krivelevich and Sudakov \[13\]. Quasi-randomness was also studied in many other settings besides graphs, such as set systems \[4\], tournaments \[5\] and hypergraphs \[6\].

The main objective of our paper is to study the quasi-randomness of graph properties given by certain graph cuts. These kind of properties were first studied by Chung and Graham in \[4, 7\]. For a real \( \alpha \in (0, 1) \), the cut property \( \mathcal{P}_C(\alpha) \) is the collection of graphs \( G \) satisfying the following: for any \( U \subseteq V(G) \) of size \( |U| = \alpha n \), we have \( e(U, V \setminus U) = p\alpha(1 - \alpha)n^2 + o(n^2) \). As it turns out, for most values of \( \alpha \), the cut property \( \mathcal{P}_C(\alpha) \) is quasi-random. In \[4, 7\], the authors proved the following beautiful theorem which characterizes the quasi-random cut properties.

**Theorem 1.2.** \( \mathcal{P}_C(\alpha) \) is quasi-random if and only if \( \alpha \neq 1/2 \).

To see that \( \mathcal{P}_C(1/2) \) is not quasi-random, Chung and Graham \[4, 7\] observed that the graph obtained by taking a random graph \( G(n/2, 2p) \) on \( n/2 \) vertices and an independent set on the remaining \( n/2 \) vertices, and then connecting these two graphs with a random bipartite graph with edge probability \( p \), satisfies \( \mathcal{P}_C(1/2) \) but is not quasi-random.

An \( r \)-cut is a partition of a vertex set \( V \) into subsets \( V_1, \ldots, V_r \), and if for a vector \( \vec{\alpha} = (\alpha_1, \ldots, \alpha_r) \), the size of the sets satisfies \( |V_i| = \alpha_i|V| \) for all \( i \), then we call this an \( \vec{\alpha} \)-cut.
An \( \vec{\alpha} \)-cut is called balanced if \( \vec{\alpha} = (1/r, \ldots, 1/r) \) for some \( r \), and is unbalanced otherwise. For a \( k \)-uniform hypergraph \( G \) and a cut \( V_1, \ldots, V_r \) of its vertex set, let \( e(V_1, \ldots, V_r) \) be the number of hyperedges which have at most one vertex in each part \( V_i \) for all \( i \).

A \( k \)-uniform hypergraph \( G \) is (weak) \( p \)-quasi-random if for every subset of vertices \( U \subset V(G) \), \( e(U) = p|U|^k + o(n^k) \). For \( r \geq k \), let \( \mathcal{P}_C(\vec{\alpha}) \) be the following: for every \( \vec{\alpha} \)-cut \( V_1, \ldots, V_r \), \( e(V_1, \ldots, V_r) = (p + o(1))n^k \sum_{S \subseteq [r], |S| = k} \prod_{i \in S} \alpha_i \). Note that previously we mentioned the example which illustrate the non-quasi-randomness of \( \mathcal{P}_C(1/2) \). As noticed by Shapira and Yuster [15], a similar construction as above shows that \( \mathcal{P}_C(1/k, \ldots, 1/k) \) is not quasi-random. In fact, they generalized Theorem 1.2 by proving the following theorem.

**Theorem 1.3.** Let \( k \geq 2 \) be a positive integer. For \( k \)-uniform hypergraphs, the cut property \( \mathcal{P}_C(\vec{\alpha}) \) is quasi-random if and only if \( \vec{\alpha} \neq (1/r, \ldots, 1/r) \) for some \( r \geq k \).

For a fixed graph \( H \), let \( \mathcal{P}_H \) be the following property: for every subset \( U \subset V \), the number of copies of \( H \) in \( U \) is \( (p|E(H)| + o(1))\left(\binom{|V|}{|V(H)|} \right) \). In [16], Simonovits and Sós proved that \( \mathcal{P}_H \) is equivalent to \( \mathcal{P}_1 \) and hence is quasi-random. For a fixed graph \( H \), as a common generalization of Chung and Graham’s and Simonovits and Sós’ theorems, we can consider the number of copies of \( H \) having one vertex in each part of a cut. Let us consider the cases when \( H \) is a clique of size \( k \).

**Definition 1.4.** Let \( k, r \) be positive integers such that \( r \geq k \geq 2 \), and let \( \vec{\alpha} = (\alpha_1, \ldots, \alpha_r) \) be a vector of positive real numbers satisfying \( \sum_{i=1}^r \alpha_i = 1 \). We say that a graph satisfies the \( K_k \) cut property \( \mathcal{C}_k(\vec{\alpha}) \) if for every \( \vec{\alpha} \)-cut \( (V_1, \ldots, V_r) \), the number of copies of \( K_k \) which have at most one vertex in each of the sets \( V_i \) is \( (p^{(\vec{\alpha})} + o(1))n^k \sum_{S \subseteq [r], |S| = k} \prod_{i \in S} \alpha_i \).

Shapira and Yuster [15] proved that for \( k \geq 3 \), \( \mathcal{C}_k(\vec{\alpha}) \) is quasi-random if \( \vec{\alpha} \) is unbalanced (note that \( \mathcal{C}_2(\vec{\alpha}) \) is quasi-random if and only if \( \vec{\alpha} \) is unbalanced). This result is a corollary of Theorem 1.3 by the following argument. For a graph \( G \) satisfying \( \mathcal{C}_k(\vec{\alpha}) \), consider the \( k \)-uniform hypergraph \( G' \) on the same vertex set where \( k \)-tuple of vertices forms an hyperedge if and only if they form a clique in \( G \). Then \( G' \) satisfies \( \mathcal{P}_C(\vec{\alpha}) \) and thus is quasi-random. By the definition of the quasi-randomness of hypergraphs, this in turn implies that the number of cliques of size \( k \) inside every subset of \( V(G) \) is “correct,” and thus by Simonovits and Sós’ result, \( G \) is quasi-random.

Note that for balanced \( \vec{\alpha} \) this approach does not give enough information, since it is not clear if there exists a graph whose hypergraph constructed by the above mentioned process is not quasi-random but satisfies \( \mathcal{P}_C(\vec{\alpha}) \) (nonetheless as the reader might suspect, the properties \( \mathcal{P}_C(\vec{\alpha}) \) and \( \mathcal{C}_k(\vec{\alpha}) \) are closely related even for balanced \( \vec{\alpha} \)). Shapira and Yuster made this observation and left the balanced case as an open question asking whether it is quasi-random or not (in fact, they asked the question for \( \vec{\alpha} = (1/k, \ldots, 1/k) \), but here we consider the slightly more general question for all balanced \( \vec{\alpha} \) as mentioned above). Janson [11] independently posed the same question in his paper that studied quasi-randomness under the framework of graph limits. In this paper, we settle this question by proving the following theorem:

**Theorem 1.5.** Fix a real \( p \in (0, 1) \) and positive integers \( r, k \) such that \( r \geq k \geq 3 \). For every positive \( \varepsilon \), there exists a positive \( \delta \) such that the following is true. If \( G \) is a graph which has density \( p \) and is \( \delta \)-close to satisfying the \( K_k \) balanced cut property \( \mathcal{C}_k(1/r, \ldots, 1/r) \), then \( G \) is \( \varepsilon \)-close to being \( p \)-quasi-random.
The rest of the paper is organized as follows. In Section 2 we introduce the notations we are going to use throughout the paper and state previously known results that we need later. In Section 3 we give a detailed proof of the most important base case of Theorem 1.5, triangle balanced cut property, i.e. $C_3(1/r, \ldots, 1/r)$. In Section 4, we prove the general case as a consequence of the base case. The last section contains some concluding remarks and open problems for further study.

2. PRELIMINARIES

Given a graph $G = (V, E)$ and two vertex sets $X, Y \subseteq V(G)$, we denote by $E(X, Y)$ the set of edges which have one end point in $X$ and the other in $Y$. Also we write $e(X, Y) = |E(X, Y)|$ to indicate the number of edges and $d(X, Y) = \frac{e(X, Y)}{|X||Y|}$ for the density. For a cut $X = (X_1, \ldots, X_r)$ of the vertex set, we say that a triangle with vertices $u, v, w$ crosses the cut $X$ if it contains at most one vertex from each set, and denote it by $(u, v, w) \cap X$. We use $Tr(X)$ for the number of triangles with vertices $(u, v, w) \cap X$. For a $k$-uniform hypergraph and a partition $V_1, \ldots, V_t$ of its vertex set $V$ into $t$ parts, we define its density vector as the vector in $\mathbb{R}^t$ indexed by the $k$-subsets of $[t]$ whose $(i_1, \ldots, i_k)$-entry is the density of hyperedges which have exactly one vertex in each of the sets $V_{i_1}, \ldots, V_{i_k}$. Throughout the paper, we always use subscripts such as $\delta_{2,6}$ to indicate that the parameter $\delta$ comes from Theorem 2.6.

To state asymptotic results, we utilize the following standard notations. For two positive-valued functions $f(n)$ and $g(n)$, write $f(n) = \Omega(g(n))$ if there exists a positive constant $c$ such that $\liminf_{n \to \infty} f(n)/g(n) \geq c$, $f(n) = o(g(n))$ if $\limsup_{n \to \infty} f(n)/g(n) = 0$. Also, $f(n) = O(g(n))$ if there exists a positive constant $C > 0$ such that $\limsup_{n \to \infty} f(n)/g(n) \leq C$.

To isolate the unnecessary complication arising from the error terms, we will use the notation $x =_{\varepsilon} y$ if $|x - y| = O(\varepsilon)$ and say that $x, y$ are $\varepsilon$-equal. For two vectors, we define $\vec{x} =_{\varepsilon} \vec{y}$ if $\|\vec{x} - \vec{y}\|_\infty = O(\varepsilon)$. We omit the proof of the following simple properties (we implicitly assume that the following operations are performed a constant number of times in total). Let $C$ and $c$ be positive constants.

1a. (Finite transitivity) If $x =_{\varepsilon} y$ and $y =_{\varepsilon} z$, then $x =_{\varepsilon} z$.
1b. (Complete transitivity) For a finite set of numbers $\{x_1, \ldots, x_n\}$, if $x_i =_{\varepsilon} x_j$ for every $i, j$, then there exists $x$ such that $x_i =_{\varepsilon} x$ for all $i$.
2. (Additivity) If $x =_{\varepsilon} z$ and $y =_{\varepsilon} w$, then $x + y =_{\varepsilon} z + w$.
3. (Scalar product) If $x =_{\varepsilon} y$ and $0 < c \leq a \leq C$, then $ax =_{\varepsilon} ay$ and $x/a =_{\varepsilon} y/a$.
4. (Product) If $x, y, z, w$ are bounded above by $C$, then $x =_{\varepsilon} y$ and $z =_{\varepsilon} w$ implies that $xz =_{\varepsilon} yw$.
5. (Square root) If both $x$ and $y$ are greater than $c$, then $x^2 =_{\varepsilon} y^2$ implies that $x =_{\varepsilon} y$.
6. For the linear equation $Ax =_{\varepsilon} \vec{y}$, if all the entries of an invertible matrix $A$ are bounded by $C$, and the determinant of $A$ is bounded from below by $c$, then $\vec{x} =_{\varepsilon} A^{-1}\vec{y}$.
7. If $xy =_{\varepsilon} 0$, then either $x =_{\sqrt{\varepsilon}} 0$ or $y =_{\sqrt{\varepsilon}} 0$.

2.1. Extremal Graph Theory

To prove the main theorem, we use the regularity lemma developed by Szemerédi [17]. Let $G = (V, E)$ be a graph and $\varepsilon > 0$ be fixed. A disjoint pair of sets $X, Y \subseteq V$ is called an $\varepsilon$-regular pair if $\forall A \subseteq X, B \subseteq Y$ such that $|A| \geq \varepsilon |X|, |B| \geq \varepsilon |X|$ satisfies $|d(X, Y) - d(A, B)| \leq \varepsilon$. A vertex partition $\{V_i\}_{i=1}^r$ is called an $\varepsilon$-regular partition if (i)
the sizes of \( V_i \) differ by at most 1, and (ii) \((V_i, V_j)\) is \( \epsilon \)-regular for all but at most \( \epsilon t^2 \) pairs \( 1 \leq i < j \leq n \). The regularity lemma states that every large enough graph admits a regular partition. In our proof, we use a slightly different form which can be found in [12]:

**Theorem 2.1 (Regularity Lemma).** For every real \( \epsilon > 0 \) and positive integers \( m, r \) there exists constants \( T(\epsilon, m) \) and \( N(\epsilon, m) \) such that given any \( n \geq N(\epsilon, m) \), the vertex set of any \( n \)-vertex graph \( G \) can be partitioned into \( t \) sets \( V_1, \ldots, V_t \) for some \( t \) divisible by \( r \) and satisfying \( m \leq t \leq T(\epsilon, m) \), so that

- \( |V_i| < [\epsilon n] \) for every \( i \).
- \( |V_i| - |V_j| \leq 1 \) for all \( i, j \).
- Construct a reduced graph \( H \) on \( t \) vertices such that \( i \sim j \) in \( H \) if and only if \((V_i, V_j)\) is \( \epsilon \)-regular in \( G \). Then the reduced graph has minimum degree at least \( (1 - \epsilon)t \).

As one can see in the following lemma, regular pairs are useful in counting small subgraphs of a graph (this lemma can easily be generalized to other subgraphs).

**Lemma 2.2.** Let \( V_1, V_2, V_3 \) be subsets of vertices. If the pair \((V_i, V_j)\) is \( \epsilon \)-regular with density \( d_{ij} \) for every distinct \( i, j \), then the number of triangles \( Tr(V_1, V_2, V_3) \) is

\[
Tr(V_1, V_2, V_3) = (d_{12}d_{23}d_{31} + O(\epsilon))|V_1||V_2||V_3|.
\]

**Proof.** If a vertex \( v \in V_1 \) has degree \( (1 + O(\epsilon))d_{12}|V_2| \) in \( V_2 \) and \( (1 + O(\epsilon))d_{13}|V_3| \) in \( V_3 \), then by the regularity of the pair \( (V_2, V_3) \), there will be \( (1 + O(\epsilon))|V_2||V_3|d_{12}d_{23}d_{31} \) triangles which contain the vertex \( v \). By the regularity of the pair \( (V_1, V_2) \), there are at least \( (1 - \epsilon)|V_1| \) vertices in \( V_1 \) which have at least \( (1 + O(\epsilon))d_{12}|V_2| \) neighbors in \( V_2 \), and similar holds for the pair \( (V_1, V_3) \). Hence there are at least \( (1 - 2\epsilon)|V_1| \) such vertices satisfying both conditions. Moreover, since each vertex in \( V_1 \) is contained in at most \( |V_2||V_3| \) triangles, there are at most \( 2\epsilon|V_1||V_2||V_3| \) triangles which do not contain such vertex from \( V_1 \). Therefore we have,

\[
Tr(V_1, V_2, V_3) = (1 + O(\epsilon))|V_1||V_2||V_3|d_{12}d_{23}d_{31} + 2\epsilon|V_1||V_2||V_3| = (d_{12}d_{23}d_{31} + O(\epsilon))|V_1||V_2||V_3|.
\]

For a fixed graph \( H \), a perfect \( H \)-factor of a large graph \( G \) is a collection of vertex disjoint copies of \( H \) that cover all the vertices of \( G \). The next theorem is a classical theorem proved by Hajnal and Szemerédi [9] which establishes a sufficient minimum degree condition for the existence of a perfect clique factor.

**Theorem 2.3 ([9]).** Let \( k \) be a fixed positive integer and \( n \) be divisible by \( k \). If \( G \) is a graph on \( n \) vertices with minimum degree at least \( (1 - 1/k)n \), then \( G \) contains a perfect \( K_k \)-factor.

### 2.2. Concentration

The following concentration result of Hoeffding [10] and Azuma [2] will be used several times during the proof (see also [14, Theorem 3.10]).
Theorem 2.4 (Hoeffding-Azuma Inequality). Let $c_1, \ldots, c_n$ be constants, and let $X_1, \ldots, X_n$ be a martingale difference sequence with $|X_k| \leq c_k$ for each $k$. Then for any $t \geq 0$,

$$
\Pr\left(\left| \sum_{i=1}^n X_i \right| \geq t \right) \leq 2 \exp\left( -\frac{t^2}{2 \sum_{i=1}^n c_i^2} \right).
$$

The next lemma is a corollary of Hoeffding-Azuma’s inequality.

Lemma 2.5. Let $G = (V, E)$ be a graph with $|V| = n$ and $|E| = d(n^2)$ for some fixed real $d$. Let $U$ be a random subset of $V$ constructed by selecting every vertex independently with probability $\alpha$. Then $e(U) = \alpha^2 d(n^2) + o(n^2)$ with probability at least $1 - e^{-O(n^{1/2})}$.

Proof: Arbitrarily label the vertices by $1, \ldots, n$ and consider the vertex exposure martingale. More precisely, let $X_k$ be the number of edges within $U$ incident to $k$ among the vertices $1, \ldots, k - 1$ ($X_k = 0$ if $k \notin U$), and note that $e(U) = X_1 + \cdots + X_n$. Also note that $Y_k = \mathbb{E}[X_1 + \cdots + X_n | X_1, \ldots, X_{k-1}]$ for $1 \leq k \leq n$ forms a martingale such that $|Y_k - Y_{k-1}| \leq n$ for all $k \geq 1$. Thus by Hoeffding-Azuma’s inequality (Theorem 2.4),

$$
\Pr(|e(U) - \mathbb{E}[e(U)]| \geq C) \leq 2e^{-2C^2/n^3}.
$$

Since $\mathbb{E}[e(U)] = \alpha^2 d(n^2)$, by selecting $C = n^{7/4}$, we obtain $e(U) = \alpha^2 d(n^2) + o(n^2)$ with probability at least $1 - e^{-O(n^{1/2})}$ (see, e.g., [1, Theorem 7.2.3] for more on vertex exposure martingales).

Note that the probability of success in this lemma can be improved by carefully choosing our parameters. However, Lemma 2.5 as stated is already strong enough for our later applications.

2.3. Quasi-Randomness of Hypergraph Cut Properties

Recall the cut property $\mathcal{P}_C(\vec{a})$ defined in the introduction, and the fact that it is closely related to the clique cut property $\mathcal{C}_k(\vec{a})$. While proving Theorem 1.3, Shapira and Yuster also characterized the structure of hypergraphs which satisfy the balanced cut property $\mathcal{P}_C(1/r, \ldots, 1/r)$. Let $p \in (0, 1)$ be fixed and $t$ be an even integer. In order to classify the $k$-uniform hypergraphs satisfying the balanced cut property, we first look at certain edge-weighted hypergraphs. Fix a set $I \subset [t]$ of size $|I| = t/2$, and consider the weighted hypergraph on the vertex set $[t]$ such that the hyperedge $e$ has density $2p|e \cap I|/k$ for all $e$. Let $u_{i,p}$ be the vector in $\mathbb{R}^{[t]}$ representing this weighted hypergraph (each coordinate corresponds to a $k$-subset of $[t]$, and the value of the vector at the coordinate is the edge-weight of that hyperedge), and let $W_{i,p}$ be the affine subspace of $\mathbb{R}^{[t]}$ spanned by the vectors $u_{i,p}$ for all possible sets $I$ of size $|I| = t/2$. In [15], the authors proved that the structure of a (non-weighted) hypergraph which is $\delta$-close to satisfying the balanced cut property $\mathcal{P}_C(1/r, \ldots, 1/r)$ can be described by the vector space $W_{i,p}$ (note that the vector which has constant weight lies in this space).
Theorem 2.6 ([15]). Let \( p \in (0, 1) \) be fixed. There exists a real \( t_0 \) such that for every \( \varepsilon > 0 \), and for every \( t \geq t_0 \) divisible by \( 2r, \) there exists \( \delta = \delta(t, \varepsilon) > 0 \) so that the following holds. If \( G \) is a \( k \)-uniform hypergraph with density \( p \) which is \( \delta \)-close to satisfying the balanced cut property \( \mathcal{P}_C(1/r, \ldots, 1/r) \), then for any partition of \( V(G) \) into \( t \) equal parts, the density vector \( \mathbf{d} \) of this partition satisfies \( \| \mathbf{d} - \mathbf{y} \|_\infty \leq \varepsilon \) for some vector \( \mathbf{y} \in W_{tp} \).

A part of the proof of Shapira and Yuster’s theorem relies on showing that certain matrices have full rank, and they establish this result by using the following famous result from algebraic combinatorics proved by Gottlieb [8]. For a finite set \( T \) and integers \( h \) and \( k \) satisfying \( |T| > h \geq k \geq 2 \), denote by \( B(T, h, k) \) the \( h \) versus \( k \) inclusion matrix of \( T \) which is the \( \binom{|T|}{h} \times \binom{|T|}{k} \) 0-1 matrix whose rows are indexed by the \( h \)-element subsets of \( T \), columns are indexed by the \( k \)-elements subsets of \( T \), and entry \((I, J)\) is 1 if and only if \( J \subset I \).

Theorem 2.7. \( \text{rank}(B(T, h, k)) = \binom{|T|}{h} \) for all \( |T| \geq h + k \).

3. BASE CASE—TRIANGLE BALANCED CUT

In this section we prove a special case, triangle balanced cut property, of the main theorem. Our proof consists of several steps. Let \( G \) be a graph which satisfies the triangle balanced cut property. First we apply the regularity lemma to describe the structure of \( G \) by an \( \varepsilon \)-regular partition \( \{V_i\}_{i=1}^t \). This step allows us to count the edges or triangles effectively using regularity of the pairs. From this point on, we focus only on the cuts whose parts consist of a union of the sets \( V_i \). In the next step, we swap some vertices of \( V_i \) and \( V_j \). By the triangle cut property, we can obtain an algebraic relation of the densities inside \( V_i \) and between \( V_i \) and \( V_j \). After doing this, the problem is transformed into solving a system of nonlinear equations, which basically implies that inside any clique of the reduced graph, most of the densities are very close to each other. Finally resorting to results from extremal graph theory, we can conclude that almost all the densities are equal and thus prove the quasi-randomness of triangle balanced cut property.

To show that our given graph is quasi-random, ideally, we would like to show that the densities of edges between pair of parts in the regular partition is (almost) equal to each other. However, instead of establishing quasi-randomness through verifying this strong condition, we will derive it from a slightly weaker condition. More specifically, we will use the fact that if in an \( \varepsilon \)-regular partition of the graph, the density of edges in most of the pairs of parts are equal to each other, then the graph is quasi-random (there are some dependencies in parameters). Following is the main theorem of this section.

Theorem 3.1. Fix a real \( p \in (0, 1) \) and an integer \( r \geq 3 \). For every positive \( \varepsilon \), there exists a positive real \( \delta \) such that the following is true. If \( G \) is a graph which has density \( p \) and is \( \delta \)-close to satisfying the triangle balanced cut property \( \mathcal{C}_3(1/r, \ldots, 1/r) \), then \( G \) is \( \varepsilon \)-close to being \( p \)-quasi-random.

Let \( G \) be a graph \( \delta \)-close to satisfying \( \mathcal{C}_3(1/r, \ldots, 1/r) \). By applying the regularity lemma, Theorem 2.1, to \( G \), we get an \( \varepsilon \)-regular equipartition \( \pi = \{V_i\}_{i=1}^t \). We can assume

---

1The authors omitted the divisibility condition in their paper [15].
that $|V_i| = \cdots = |V_r|$ by deleting at most $t$ vertices. The reason this can be done is that later when we use the triangle cut property to count the number of triangles, the error term that this deletion creates is at most $mn^2$ which is negligible comparing to $\delta n^3$ when $n$ is sufficiently large. Also in the definition of quasi-randomness, the error term from counting edges is at most $m^2$, which is also $o(n^2)$.

Now denote the edge density within $V_i$ by $x_i$, the edge density between $V_i$ and $V_j$ by $d_{ij}$, and the density of triangles in the tripartite graph formed by $(V_i, V_j, V_k)$ by $d_{ijk}$. Call a triple $(V_i, V_j, V_k)$ regular if each of the three pairs is regular.

Consider a family $\{\pi_{\alpha}\}_{\alpha \in [0,1]}$ of partitions of $G$ given as follows:

$$\pi_{\alpha} = ((1-\alpha)V_1 + \alpha V_2, \alpha V_1 + (1-\alpha)V_2, V_3, \ldots, V_r).$$

In other words, we pick $U_1$ and $U_2$ both containing $\alpha$-proportion of vertices in $V_1$ and $V_2$ uniformly at random and exchange them to form a new equipartition $\pi_{\alpha}$. To be precise, for fixed $\alpha$, the notation $\pi_{\alpha}$ represents a family of random partitions and not necessarily an individual partition. For convenience we assume that $\pi_{\alpha}$ is a partition constructed as above which satisfies some explicit properties that we soon mention which a.a.s. hold for random partitions. Denote the new triangle density vector of $\pi_{\alpha}$ by $\mathbf{d}^\alpha = (d_{ijk}^\alpha)$.

Note that every $(1/r, \ldots, 1/r)$-cut $X = \{X_1, \ldots, X_r\}$ of the index set $[t]$ also gives a $(1/r, \ldots, 1/r)$-cut of $V(G)$. With a slight abuse of notation, we use $(i,j,k) \cap X$ to indicate that $V_i$, $V_j$ and $V_k$ completely belong to different parts of the cut induced by $X$.

By the triangle balanced cut property, for every $\alpha \in [0,1]$,

$$\left(p^3 \pm \delta\right) \left(\frac{n}{r}\right)^3 \cdot \left(\frac{r}{3}\right) = \sum_{(i,j,k) \cap X} Tr(V_i, V_j, V_k) = \sum_{(i,j,k) \cap X} d_{ijk}^\alpha \left(\frac{n}{r}\right)^3.$$  

So $\sum_{(i,j,k) \cap X} d_{ijk}^\alpha = (p^3 \pm \delta) \left(\frac{r}{3}\right)^3$. Let $M$ be the $\left( \frac{r}{3}\right) \times \left( \frac{r}{3}\right)$ 0-1 matrix whose rows are indexed by the $(1/r, \ldots, 1/r)$-cuts of the vertex set $[t]$ and columns are indexed by the triples $(i,j,k)$, where the $(X, (i,j,k))$-entry of $M$ is 1 if and only if $(i,j,k) \cap X$. The observation above implies $Md^\alpha = (p^3 \pm \delta) \left(\frac{r}{3}\right)^3 \cdot 1$ where $1$ is the all-one vector. Thus if we let $d' = d^{1/2} - \frac{1}{3}d^0 - \frac{1}{3}d^1$, then $Md' = \alpha^\alpha \cdot 0$. From this equation we hope to get useful information about the densities $x_i$ and $d_{ij}$. With the help of the following lemma, we can compute the new densities $d_{ijk}^\alpha$, and thus the modified density vector $d'$, in terms of the densities $x_i$ and $d_{ij}$.

**Lemma 3.2.** Let $\varepsilon$ satisfy $0 < \varepsilon < d_{ij}/2$ for every $i,j$ and assume that the graph $G$ is large enough. Then for all $\alpha \in (\varepsilon, 1-\varepsilon)$, there exists a choice of sets $U_1, U_2$ such that the following holds.

1. $d_{ijk}^\alpha = \begin{cases} d_{ijk} & \text{if } (i,j,k) \cap \{1,2\} = \emptyset \\ (1-\alpha)d_{ijk} + \alpha d_{jk} + o(1) & \text{if } i = 1 \text{ and } j \notin \{j,k\} \\ \alpha d_{ijk} + (1-\alpha)d_{jk} + o(1) & \text{if } i = 2 \text{ and } j \notin \{j,k\} \\ \text{see (2)} & \text{if } i = 1 \text{ and } j = 2 \end{cases}$

2. If $(V_1, V_2, V_k)$ is a regular triple, then

$$d_{12k}^\alpha = ((1-\alpha)^2 + \alpha^2)d_{12}d_{1k}d_{2k} + \alpha(1-\alpha)(x_1d_{1k}^2 + x_2d_{2k}^2) + O(\varepsilon).$$

*Random Structures and Algorithms* DOI 10.1002/rsa
Let \( d'_{ijk} = d''_{ijk} - (1 - \alpha)d''_{ijk} - \alpha d''_{ijk} \). Then

\[
d'_{ijk} = \begin{cases} 
0 & \text{if } \{i, j, k\} \cap \{1, 2\} = \emptyset \\
\alpha(1) & \text{if } i = 1 \text{ and } 2 \notin \{j, k\} \\
\alpha(1) & \text{if } i = 2 \text{ and } 1 \notin \{j, k\}
\end{cases}
\]

Moreover, for the case \( i = 1 \) and \( j = 2 \), if \((V_1, V_2, V_k)\) is a regular triple, then

\[
d'_{12k} = \alpha(1 - \alpha)(x_1d''_{1k} + x_2d''_{2k} - 2d_{12}d_{1k}d_{2k}) + O(\varepsilon).
\]

**Proof.** Throughout the proof, we rely on the fact that some events hold with probability \( 1 - o(1) \). Since there are fixed number of events involved, without further mentioning, we will assume that all the involved events happen together at the same time.

(1) The claim clearly holds for the cases \( \alpha = 0 \) and \( \alpha = 1 \).

For \( \alpha \in (0, 1) \), if \( \{i, j, k\} \cap \{1, 2\} = \emptyset \), then the density \( d''_{ijk} \) is not affected by the swap of vertices in \( V_1 \) and \( V_2 \) so it remains the same with \( d'_{ijk} \). In the case that \( \{i, j, k\} \cap \{1, 2\} = \{1\} \), without loss of generality we assume \( i = 1 \) and \( j, k \neq 2 \). We also assume that there are \( S_i \) triangles with a fixed vertex \( x \in V_1 \cup V_2 \) and two other vertices belonging to \( V_j \) and \( V_k \) respectively (note that \( S_i \leq |V_j||V_k| \)). After swapping subset \( U_1 \subset V_1 \) with \( U_2 \subset V_2 \) such that \( |U_1| = |U_2| = \alpha|V_i| \), we know that the number of triangles in triple \( (((V_1 \cup U_2) \setminus U_1), V_j, V_k) \) changes by \( \sum_{u \in U_2} S_u - \sum_{u \in U_1} S_u \).

Assume \( |V_i| = m \), instead of taking \( \alpha m \) vertices uniformly at random, take every vertex in \( V_1 \) (or \( V_2 \)) independently with probability \( \alpha \). This gives random variables \( X_i \) for \( 1 \leq i \leq m \) having Bernoulli distribution with parameter \( \alpha \). Let \( R = \sum_{i=1}^m X_i \) and \( S = \sum_{i=1}^m X_i S_i \). These random variables represent the number of vertices chosen for \( U_1 \), and the number of triangles in the triple that contain these chosen vertices, respectively. It is easy to see

\[
Pr(R = \alpha m) = \alpha^m (1 - \alpha)^{(1 - \alpha)m} \left(\frac{m}{\alpha m}\right) \sim \Omega \left(\frac{1}{\sqrt{\alpha(1 - \alpha)}} m^{-1/2}\right),
\]

and by Hoeffding-Azuma’s inequality (Theorem 2.4)

\[
Pr(|S - \mathbb{E}S| \geq C) \leq 2 \exp \left( -\frac{C^2}{2 \sum_{i=1}^m S_i^2} \right) \leq 2 \exp \left( -\frac{C^2}{2m^3} \right).
\]

Let \( C = m^2 \), and the second probability decreases much faster than the first probability, thus we know that conditioned on the event \( R = \alpha m \), \( S \) is also concentrated at its expectation \( \mathbb{E}S = \sum_{i=1}^m \alpha S_i = \alpha d_{ijk} m^3 \). From here we know the number of triangles changes by

\[
\sum_{u \in U_2} S_u - \sum_{u \in U_1} S_u = \alpha d_{2jk} m^3 - \alpha d_{ijk} m^3 + o(m^3).
\]

Therefore the new density is

\[
d''_{1jk} = d_{1jk} + \left( \sum_{u \in U_2} S_u - \sum_{u \in U_1} S_u \right) / m^3 = (1 - \alpha)d_{1jk} + \alpha d_{2jk} + o(1).
\]

We can use a similar method to compute \( d''_{2jk} \) when \( 1 \notin \{j, k\} \).
(2) Let \( U_1 \subset V_1, U_2 \subset V_2 \) be as in (1), and let \( V'_1 = (V_1 \setminus U_1) \cup U_2, V'_2 = (V_2 \setminus U_2) \cup U_1 \). Then we have the identity
\[
Tr(V'_1, V'_2, V_k) = Tr(U_1, U_2, V_k) + Tr(U_1, V_1 \setminus U_1, V_k) + Tr(V_2 \setminus U_2, U_2, V_k) + Tr(V_2 \setminus U_2, V_1 \setminus U_1, V_k).
\]

Since \( \alpha \in (\varepsilon, 1 - \varepsilon) \), the triples \((U_1, U_2, V_k)\) and \((V_1 \setminus U_1, V_2 \setminus U_2, V_k)\) are regular. Thus by Lemma 2.2,
\[
Tr(U_1, U_2, V_k) = (d_{12}d_{1k}d_{2k} + O(\varepsilon))|U_1||U_2||V_k| = (d_{12}d_{1k}d_{2k} + O(\varepsilon))\alpha^2 m^3 \quad \text{and,}
\]
\[
Tr(V_1 \setminus U_1, V_2 \setminus U_2, V_k) = (d_{12}d_{1k}d_{2k} + O(\varepsilon))(1 - \alpha)^2 m^3.
\]

To compute \( Tr(U_1, V_1 \setminus U_1, V_k) \), let \( E^k_1 \subset E(V_1) \) be the collection of edges such that their end points have \((d_{ik} \pm \varepsilon)^2 m\) common neighbors in \( V_k \). By the regularity of the pair \((V_1, V_k)\), there are at most \(2\varepsilon m\) vertices in \( V_1 \) which do not have \((d_{ik} \pm \varepsilon)^m\) neighbors in \( V_k \), otherwise taking this set of vertices and \( V_k \) will contradict the regularity. If \( v \) is not such a vertex, then \( \varepsilon < d_{ik} - \varepsilon \) by the hypothesis \( \varepsilon < d_{ij}/2 \), by using the regularity, we see that there are at most \(2\varepsilon m\) other vertices in \( V_1 \) which do not have \((d_{ik} \pm \varepsilon)^2 m\) common neighbors with \( v \). Consequently there are at most \(4\varepsilon m^2\) edges inside \( V_1 \) which do not have \((d_{ik} \pm \varepsilon)^2 m\) common neighbors inside \( V_k \). We call these edges “exceptional”. Recall that \( x_1 \) denotes the density of edges in \( V_1 \), thus \(|E^k_1| \geq x_1(m^2) - 4\varepsilon m^2\). By Lemma 2.5 and the calculation from part (1) there exists a choice of \( U_1 \) of size \( \alpha m \) such that,
\[
|E^k_1(U_1, V_1 \setminus U_1)| = |E^k_1| - |E^k_1(U_1)| - |E^k_1(V_1 \setminus U_1)| = (1 - \alpha^2 - (1 - \alpha)^2 + o(1))|E^k_1| = \alpha(1 - \alpha)x_1 m^2 + O(\varepsilon)m^2.
\]

for all \( k \). Note that the number of triangles \( Tr(U_1, V_1 \setminus U_1, V_k) \) can be computed by adding the number of triangles containing the edges in \( E^k_1 \) and then the number of triangles containing the “exceptional” edges (recall that there are at most \(O(\varepsilon)m^2 \) of the such edges). The latter can be crudely bounded by \(O(\varepsilon)m^2 \cdot m \leq O(\varepsilon)m^3\). Since each edge in \( E^k_1 \) is contained in \((d_{ik} \pm \varepsilon)^2 m\) triangles (within the triple \((V_1, V_2, V_k)\)),
\[
Tr(U_1, V_1 \setminus U_1, V_k) = |E^k_1(U_1, V_1 \setminus U_1)| \cdot (d_{1k}^2 + O(\varepsilon))m + O(\varepsilon)m^3 = \alpha(1 - \alpha)x_1 d_{1k}^2 m^3 + O(\varepsilon)m^3.
\]

Similarly we can show that there exists a choice of \( U_2 \) of size \( \alpha m \) that gives
\[
Tr(U_2, V_2 \setminus U_2, V_k) = \alpha(1 - \alpha)x_2 d_{2k}^2 m^3 + O(\varepsilon)m^3.
\]

for all \( k \). Combining all the results together, we can conclude the existence of sets \( U_1, U_2 \) such that
\[
d_{12} = ((1 - \alpha)^2 + \alpha^2)d_{12}d_{1k}d_{2k} + \alpha(1 - \alpha)(x_1 d_{1k}^2 + x_2 d_{2k}^2) + O(\varepsilon).
\]

Part (3) is just a straightforward computation from the definition of \( d'_{jk} \) and (2).

**Lemma 3.3.** Let \( \varepsilon \) satisfy \( 0 < \varepsilon < d_{ij}/2 \) for every \( i, j \) and assume that the graph \( G \) is large enough. If \((V_i, V_j, V_k)\) is a regular triple, then \( x_i d_{jk}^2 + x_j d_{ik}^2 - 2d_{ij}d_{ik}d_{jk} = \delta^3 + \varepsilon 0 \).

*Random Structures and Algorithms DOI 10.1002rsa*
Proof. As mentioned before Lemma 3.2, the vector \( d' = d^{1/2} - \frac{1}{2}d^0 - \frac{1}{2}d^1 \) satisfies \( Md' = \delta_3 0 \). For an index \( k \neq 1, 2 \), consider a balanced partition \( \mathbf{X} \) of the vertex set \([t]\) such that 1 and 2 lie in different parts, and let \( \mathbf{Y} \) be the union of the parts which contains neither 1 nor 2. Then by Lemma 3.2 (3),

\[
0 = \delta_3 \sum_{(i,j,k) \notin \mathbf{X}} d'_{ijk} = t^3 \cdot o(1) + \sum_{k \in \mathbf{Y}} d'_{12k},
\]

where \( o(1) \) goes to 0 as the number of vertices in the graph \( G \) grows. Since \( \mathbf{Y} \) can be an arbitrary set of size \((r - 2)t/r\) not containing 1 and 2, this immediately implies that \( d'_{12k} = \delta_3 0 \) for all \( k \). Thus if \((V_1, V_2, V_k)\) is a regular triple, then by Lemma 3.2 (3), \( x_1d_{1k}^2 + x_2d_{2k}^2 - 2d_{12}d_{1k}d_{2k} = \delta_3 0 \). By symmetry, we can replace 1 and 2 by arbitrary indices \( i, j \).

Using Theorem 2.6 which characterizes the non quasi-random hypergraphs satisfying the balanced cut property, we can prove the following lemma which allows us to bound the densities from below.

**Lemma 3.4.** There exists \( t_0 \) such that for fixed \( p \in (0, 1) \) and every \( t \geq t_0 \) which is divisible by 2\( r \), there exist \( c = c(p) \) and \( \delta_0 = \delta_0(t, p) > 0 \) so that the following holds for every \( \delta \leq \delta_0 \). If \( G \) is a graph with density \( p \) which is \( \delta \)-close to satisfying the triangle balanced cut property, then for any partition \( \pi \) of \( V(G) \) into \( t \) equal parts, the density vector \( \mathbf{d} = (d_{ij})_{i,j} \) satisfies \( d_{ij} \geq c \) for all distinct \( i, j \in [t] \).

Proof. Let \( t_0 = t_{2r}, \varepsilon = p^3/8 \), and for a given \( t \geq t_0 \) divisible by 2\( r \), let \( \delta_0 = \min\{\delta_{2r}(t, \varepsilon), p^3/10\} \). Let \( V = V(G) \), and let \( G' \) be the hypergraph over the vertex set \( V \) such that \( (i,j,k) \in E(G') \) if and only if \( i,j,k \) forms a triangle in the graph \( G \). Let \( \pi \) be an arbitrary partition of \( V \) into \( t \) equal parts \( V_1, \ldots, V_t \), and let \((d_{ijk})_{i,j,k}\) be the density vector of the graph \( G \), and \((d_{ijk})_{i,j,k}\) be the density vector of the hypergraph \( G' \) with respect to \( \pi \). It suffices to show the bound \( d_{ij} \geq p^3/10 \) for every distinct \( i,j \in [t] \). For simplicity we will only verify it for \( d_{12} \). Note that the number of triangles which cross \( V_1, V_2, V_k \) is at most \( e(V_1, V_2) \cdot |V_k| = (|V_1||V_2|d_{12}) \cdot |V_k| \), and thus \( d_{12k} \leq d_{12} \) for all \( k \geq 3 \). Consequently, by summing it up over all choices of \( k \), we obtain the following inequality which will be crucial in our argument:

\[
\sum_{k=3}^{t} d_{12k} \leq (t-2) \cdot d_{12}. \tag{1}
\]

Since \( G \) is \( \delta \)-close to satisfying the triangle balanced cut property, we know that the density \( q \) of triangles is at least \( q \geq p^3 - \delta \). By Theorem 2.6, \((d_{ijk})_{i,j,k}\) is \( \varepsilon \)-equal to some vector in \( W_{t,q} \). Recall that the vectors in \( W_{t,q} \) can be expressed as an affine combination of the vectors \( u_{\mathbf{X},I} = (u_{ijk}')_{i,j,k} \) for sets \( I \subset [t] \) of size \(|I| = t/2\), and note that the following is true no matter how we choose the set \( I \) (recall that by definition we have \( u_{12k}' = \frac{2q}{3}|I \cap \{1, 2, k\}|\)):

\[
\sum_{k=3}^{t} u_{12k}' \geq \sum_{k \in I \setminus \{1, 2\}} \frac{2q}{3} \geq \left( \frac{t}{2} - 2 \right) \frac{2q}{3}.
\]

Since \((d_{ijk})_{i,j,k}\) is \( \varepsilon \)-equal to an affine combination of these vectors and an arbitrary affine combination still satisfies the inequality above, for large enough \( t \) and sufficiently small \( \varepsilon \) we have

Random Structures and Algorithms DOI 10.1002/rsa
\[
\sum_{k=3}^{t} d_{12k} \geq \left( \frac{t}{2} - 2 \right) \frac{2q}{3} - \epsilon t \geq \frac{tq}{3} - \frac{4q}{6} - \frac{tq}{6} \geq \frac{tq}{8}.
\] (2)

By combining (1) and (2), we obtain \( d_{12} \geq q/8 \geq (p^3 - \delta)/8 \geq p^3/10 \). Similarly we can deduce \( d_{ij} \geq p^3/10 \) for all distinct \( i, j \in [t] \).

Since Lemma 3.4 asserts that all the pairwise densities \( d_{ij} \) are bounded from below by some constant, we are allowed to divide each side of an \( \epsilon \)-equality by \( d_{ij} \). This turns out to be a crucial ingredient in solving the equations given by Lemma 3.3.

**Lemma 3.5.** Given a positive real \( c \) and an integer \( n \geq 4 \), if \( x_i \geq 0 \) for every \( i \), \( d_{ij} \geq c \) for every distinct \( i, j \in [n] \), and \( x_i d_{12}^2 + x_j d_{23}^2 - 2d_{ij} d_{ik} d_{jk} = \epsilon \) 0 for every distinct \( i, j, k \in [n] \), then there exists \( s \in [n] \), \( x, y > 0 \) such that for any distinct \( i, j \neq s \), we have \( d_{ij} = \epsilon \sqrt{x} \) and, for any \( i \neq s \), \( d_{is} = \epsilon \sqrt{y} \). Moreover, \( x_i = \epsilon \sqrt{x} \) if \( i \neq s \) and \( x_s = \epsilon \sqrt{y} (2y - x) \) (see, Fig. 1).

**Proof.** Throughout the proof, we heavily rely on the properties of \( \epsilon \)-equality given in Section 2.

First consider the case \( n = 4 \). By taking \((i, j, k) = (1, 2, 3), (2, 3, 1), (3, 1, 2)\) respectively, we get the following system of equations:

\[
\begin{align*}
    d_{13}^2 x_1 + d_{23}^2 x_2 &= \epsilon \quad 2d_{12} d_{13} d_{23} \\
    d_{12}^2 x_1 + d_{23}^2 x_3 &= \epsilon \quad 2d_{12} d_{13} d_{23} \\
    d_{12}^2 x_2 + d_{13}^2 x_3 &= \epsilon \quad 2d_{12} d_{13} d_{23}
\end{align*}
\] (3)

Considering this as a system of linear equations with unknowns \( x_1, x_2, x_3 \), the determinant of the coefficient matrix becomes \( 2d_{12}^2 d_{13}^2 d_{23} \geq 2c^4 \). Moreover, the coefficients in the matrix

![Fig. 1. Structure of solution for \( n = 4 \) when \( z = \epsilon \), with vertices 1 and 2 permuted if \( z = \epsilon \).](Random Structures and Algorithms DOI 10.1002/rsa)
are bounded from above by 1. Therefore we can solve the linear system by appealing to property (6) of ε-equality and get

\[
\begin{align*}
    x_1 &= \varepsilon \frac{d_{23}}{d_{12}d_{13}} (d_{12}^2 + d_{13}^2 - d_{23}^2) \\
    x_2 &= \varepsilon \frac{d_{13}}{d_{12}d_{23}} (d_{12}^2 + d_{23}^2 - d_{13}^2) \\
    x_3 &= \varepsilon \frac{d_{12}}{d_{13}d_{23}} (d_{13}^2 + d_{23}^2 - d_{12}^2)
\end{align*}
\]

(4)

Then

\[
x_1 x_2 = \varepsilon \frac{d_{23}}{d_{12}d_{13}} (d_{12}^2 + d_{13}^2 - d_{23}^2) \cdot \frac{d_{13}}{d_{12}d_{23}} (d_{12}^2 + d_{23}^2 - d_{13}^2) = \varepsilon \frac{1}{d_{12}^4} \left[ d_{12}^4 - (d_{13}^2 - d_{23}^2)^2 \right] = \varepsilon \frac{1}{d_{12}^2} \left[ d_{12}^4 - (d_{14}^2 - d_{24}^2)^2 \right].
\]

(5)

The last equation comes from repeating the same step for the system of equations for indices 1, 2, and 4. Equation (5) implies \(d_{13}^2 - d_{23}^2 = \varepsilon (d_{14}^2 - d_{24}^2)\), and \(d_{13}^2 - d_{23}^2 = \varepsilon (d_{14}^2 - d_{24}^2)\) for all distinct \(i, j, k, l\) in general. Assume that there exists an assignment \(\{i, j, k, l\} = \{1, 2, 3, 4\}\) such that \(d_{ik}^2 - d_{jk}^2 = \varepsilon d_{il}^2 - d_{jl}^2 \not= 0\) (we call such case as a “flip”). Without loss of generality let \(d_{13}^2 - d_{23}^2 = \varepsilon (d_{14}^2 - d_{24}^2)\). By Eq. (3), we know that \(x_1 d_{12}^2 + x_3 d_{13}^2 = \varepsilon 2d_{12}d_{13}d_{23} = \varepsilon x_2 d_{12}^2 + x_3 d_{13}^2\), from which we get

\[d_{12}^2 (x_1 - x_2) = \varepsilon (d_{13}^2 - d_{23}^2) x_3.\]

Replace the index 3 by 4 and we get

\[d_{12}^2 (x_1 - x_2) = \varepsilon (d_{14}^2 - d_{24}^2) x_4.\]

By the assumption on a “flip”, by subtracting the two equalities we get \(x_3 + x_4 = \varepsilon 0\), thus \(x_3 = \varepsilon 0\) and \(x_4 = \varepsilon 0\) by their nonnegativity. This is impossible from the equation \(x_3 d_{11}^2 + x_4 d_{14}^2 = \varepsilon 2d_{13}d_{14}d_{34}\) and the fact \(d_{ij} \geq c\). Therefore no flip exists and we have

\[
d_{ik}^2 - d_{jk}^2 = \varepsilon d_{il}^2 - d_{jl}^2 \quad \forall (i, j, k, l) = \{1, 2, 3, 4\}.
\]

(6)

Since \(d_{ij} \geq c\), the sum of \(d_{12}^2 + d_{13}^2 - d_{23}^2\), \(d_{12}^2 + d_{13}^2 - d_{23}^2\), \(d_{13}^2 + d_{14}^2 - d_{23}^2\) and \(d_{13}^2 + d_{14}^2 - d_{23}^2\) is equal to \(d_{12}^2 + d_{13}^2 + d_{23}^2 \geq 3c^2\). So at least one of the terms is greater than \(c^2\) without loss of generality we can assume \(d_{12}^2 + d_{13}^2 - d_{23}^2 \geq c^2\). Recall that \(x_1 = \varepsilon \frac{d_{23}}{d_{13}d_{12}} (d_{12}^2 + d_{13}^2 - d_{23}^2)\).

By Eq. (6), we also have \(x_1 = \varepsilon \frac{d_{24}}{d_{14}d_{12}} (d_{12}^2 + d_{14}^2 - d_{24}^2) = \varepsilon \frac{d_{24}}{d_{14}d_{12}} (d_{12}^2 + d_{14}^2 - d_{23}^2)\). Therefore

\[
\frac{d_{23}}{d_{13}} = \varepsilon \frac{d_{24}}{d_{14}}.
\]

By appealing to the bound \(d_{ij} \geq c\), we get \(d_{23}d_{14} = \varepsilon d_{24}d_{13}\) and \(d_{23}^2 d_{14} = \varepsilon d_{24}^2 d_{13}\), which implies

\[
(d_{13}^2 - d_{23}^2)(d_{14}^2 - d_{24}^2) = d_{13}^2 (d_{14}^2 - d_{24}^2) - d_{13}^2 d_{14}^2 + d_{23}^2 d_{14}^2 = \varepsilon d_{13}^2 (d_{14}^2 - d_{24}^2) - d_{13}^2 d_{14}^2 + d_{13}^2 d_{24}^2 = 0.
\]
So either \( d_{13}^2 = \sqrt{\tau} \) or \( d_{14}^2 = \sqrt{\tau} \). Thus at this point we may assume the existence of indices \( i,j,k \) satisfying \( d_{ik}^2 = \sqrt{\tau} \). Assume that \( d_{13}^2 = \sqrt{\tau} \) as the other case can be handled identically.

So \( d_{13}^2 = \sqrt{\tau} x \), \( d_{14}^2 = \sqrt{\tau} y \) for some \( x \) and by Eq. (6) we have \( d_{23}^2 = \sqrt{\tau} y \), \( d_{24}^2 = \sqrt{\tau} y \) for some \( y \). We let \( d_{12}^2 = z \), and the equation \( d_{14}^2 - d_{12}^2 = d_{14}^2 - d_{24}^2 \) given by (6) translates to \( d_{12}^2 = x + y - z \). Moreover, from Eq. (4) for indices \( \{1, 3, 4\} \) and \( \{1, 2, 4\} \) we know that

\[
x_1 = \frac{d_{34}}{d_{14}d_{13}} (d_{14}^2 + d_{13}^2 - d_{34}^2) = \frac{d_{24}}{d_{14}d_{12}} (d_{14}^2 + d_{12}^2 - d_{24}^2).
\]

If we plug all equalities for \( d_{ij} \) into this identity, we get

\[
(2x - z) \sqrt{x} = \sqrt{\tau} (2x - z) \frac{\sqrt{y}}{\sqrt{x(y + z)}},
\]

So either \( \frac{\sqrt{x}}{x} = \frac{\sqrt{y}}{\sqrt{z}} \) or \( z = x^{1/4} \). In the first case, by solving this equation we get either \( z = x^{1/4} \) or \( z = y^{1/4} \) (before multiplying each side of the equation by its denominators, one must establish the fact that \( x + y - z \) is bounded away from 0. This can be done by first noting that equation \( \frac{x}{z} = \frac{y}{z} \) holds, and then realizing that the left hand side is bounded from above). Both of the above solutions gives us values for \( x \)'s and \( y \)'s as claimed (see Fig. 1, for the case \( z = x^{1/4} \))

In the second case \( z = y^{1/4} \), we consider the equation

\[
x_2 = \frac{d_{34}}{d_{24}d_{23}} (d_{24}^2 + d_{23}^2 - d_{34}^2) = \frac{d_{14}}{d_{24}d_{12}} (d_{24}^2 + d_{12}^2 - d_{14}^2)
\]

to get

\[
(2y - z) \sqrt{y} = \sqrt{\tau} (2y - z) \frac{\sqrt{\tau}}{\sqrt{x(y + z)}}.
\]

By the previous analysis we may assume \( z = 2y \), which implies \( x = y^{1/4} \) and \( d_{12}^2 = x^{1/4} \). This is impossible by the fact \( d_{12} \geq c \).

Note that we have studied the case \( n = 4 \). For \( n = 5 \), suppose not all the edge densities are \( \tau \)-equal to the same value. In this case, there must be four vertices such that not all the densities between them are equal. Without loss of generality, \( d_{12} = \tau \), \( d_{13} = \tau \), \( d_{14} = \tau \), \( d_{15} \) should be called \( x \), and \( d_{23} = \tau \), \( d_{24} = \tau \), \( d_{25} = \tau \) should be called \( y \), and \( x \neq y \). Now let us consider the collections of vertices \( \{v_1, v_2, v_3, v_5\} \), \( \{v_1, v_2, v_4, v_5\} \), \( \{v_1, v_3, v_4, v_5\} \). From the case \( n = 4 \), we know that \( d_{15} = \tau x \), \( d_{25} = \tau d_{35} = \tau d_{45} = \tau y \). By repeating this process we can generalize it to arbitrary \( n \geq 5 \).

Note that if in the regular partition, every pair of sets were regular, then Lemma 3.5 itself forces the graph to be quasi-random, as apart from one part (which is negligible), all the densities are equal. However, the regularity lemma inevitably produces a partition which contains some irregular pairs, and in the remainder of the proof of Theorem 3.1 we will show how to handle this subtlety. The main idea is that since there are only a small number of irregular pairs, the reduced graph will contain many cliques, and thus that we can use Lemma 3.5 to study its structure.

From now on in the reduced graph, when a clique of size at least 4 is given, we will call the exceptional vertex \( s \) “bad” and all others “good” vertices. We also call the densities
an equitable partition of the vertices of \( H \) and

| 138 |

Random Structures and Algorithms

| 138 |

138

Now we can combine Lemmas 3.3, 3.4, and 3.5 above to prove the main theorem which says that the triangle balanced cut property is quasi-random.

**Proof of Theorem 3.1 (triangle case).** Let \( c = c_{3,4}(p) \). We may assume that \( \varepsilon < \min(c/2, 1/32) \). Let \( t_0 = t_{3,4} \) and \( T = T_{2,1}(\varepsilon, t_0) \). Let \( \delta = \min_{0 \leq t \leq T} (\varepsilon/t^2, \delta_{3,4}(t, p)) \).

Let \( G \) be a graph which is \( \delta \)-close to satisfying \( C_3(1/r, \ldots, 1/r) \). Consider the \( \varepsilon \)-regular equipartition \( \pi \) of \( G: V(G) = V_1 \cup \cdots \cup V_r \) we mentioned before. This gives a reduced graph \( H \) on \( t \) vertices of minimum degree at least \((1 - \varepsilon)t\) (we may assume that \( t \) is divisible by \( 4r \)). Every edge \( ij \) corresponds to an \( \varepsilon \)-regular pair \((V_i, V_j)\). We mark on each edge of \( H \) a weight \( d_{ij} \) which is the density of edges in \((V_i, V_j)\), and also the density \( x_i \) inside \( V_i \) on the vertices. Parameters are chosen so that Lemma 3.3 and Lemma 3.4 holds. Moreover, by the fact \( \delta t^2 \leq \varepsilon \), we have \( x_i d_{ij}^2 + x_j d_{jk}^2 - 2d_{ij}d_{ik}d_{jk} = 0 \) for every regular triple \((V_i, V_j, V_k)\). Thus whenever there is a clique of size at least 4 in \( H \), by Lemma 3.5 we know that all the densities are \( \varepsilon \)-equal to each other, except for at most one “bad” vertex. Since \( \varepsilon < 1/32 \) and \( 4t \), we can apply Hajnal-Szemerédi theorem (Theorem 2.3) to the reduced graph \( H \) and get an equitable partition of the vertices of \( H \) into vertex disjoint 4-cliques \( C_1, \ldots, C_{4t} \).

For every 4-clique \( C_i \), from Lemma 3.5 we know that there is at most one “bad” vertex. For two 4-cliques \( C_i \) and \( C_j \), we can consider the bipartite graph \( B(C_i, C_j) \) between them which is induced from \( H \). If \( B(C_i, C_j) = K_{4,4} \), then it contains a subgraph isomorphic to \( K_{2,2} \) where the vertices are “good” (two vertices are good in \( C_i \) and other two in \( C_j \)). If we apply the structural lemma, Lemma 3.5, to this new 4-clique (together with two edges coming from the two known cliques), we get that the “good” densities of \( C_i \) and \( C_j \) are \( \varepsilon \)-equal to each other.

Now consider the reduced graph \( H' \) whose vertices correspond to the 4-cliques \( C_i \) and \( C_j \) are adjacent in \( H' \) if and only if there is a complete bipartite graph between them. It is easy to see that the minimum degree \( \delta(H') \geq (1 - 16\varepsilon)|H'| \), since for any clique \( C_i \), there are at most \( 4\varepsilon|H'| \) edges intersecting it which is not in \( H \), therefore \( \delta(H') \geq |H'| - 4\varepsilon|H'| = (1 - 16\varepsilon)|H'| \). Take any two vertices \( u', v' \in V(H') \), then \( d(u') + d(v') \geq (1 - 16\varepsilon)|H'| > |H'| \) for \( \varepsilon < 1/32 \), they have a common neighbor \( w' \), and thus by the discussion above, the “good” density in \( C_{u'} \) or \( C_{v'} \) are \( \varepsilon \)-equal to the “good” density in \( C_{w'} \). So all the “good” densities are \( \varepsilon \)-equal to each other. Thus by the total transitivity of \( \varepsilon \)-equality (see, Section 2), all the “good” densities are \( \varepsilon \)-equal to \( p' \) for some \( p' \).

We would like to show that \( d_{ij} = d_{ij} \) for all but at most \( O(\varepsilon) t^2 \) edges of the reduced graph \( H \). We already verified this for “good” edges \( \{i, j\} \) belonging to the cliques \( C_1, \ldots, C_{4t} \). If \( C_i \) is adjacent to \( C_j \) in \( H' \) then they actually form a clique of size 8 in \( H \), and by Lemma 3.5 there is at most one “bad” vertex there. Hence there is at most one “bad vertex” in two adjacent cliques. Consequently the total number of cliques that contain at least one “bad” vertex cannot exceed the independence number of \( H' \), which is at most \( |H'| - \delta(H') \leq 16\varepsilon t \). Hence among the cliques \( C_1, \ldots, C_{4t} \) there are at most \( 16t \) cliques which contain at least one “bad” vertex. Moreover, the density of an edge in \( H \) which is part of a \( K_{4,4} \) connecting two “good” cliques \( C_i \) and \( C_j \) are \( \varepsilon \)-equal to \( p' \) again by Lemma 3.5. Among the remaining edges, all but at most \( 16\varepsilon t^2 \) are such edges connecting two “good” cliques with a \( K_{4,4} \) as otherwise \( e(H) < \binom{|H|}{2} - \varepsilon t^2 \) which is a contradiction. Therefore all but at most \( O(\varepsilon) t^2 \) edges of \( H \) have density \( \varepsilon \)-equal to \( p' \). This in turn implies that the density of \( G \) is equal to \( p' + O(\varepsilon) \). On the other hand we know that the density is \( p' \), thus \( p' = p \).
Now by verifying that $G$ satisfies $\mathcal{P}_2(1/2)$ (see Theorem 1.1), we will show that $G$ is quasi-random. For an arbitrary subset $U \subset V(G)$ of size $n/2$, let us compute the number of edges in $e(U)$ and estimate its difference with the number of edges of a subset of size $n/2$ in $G(n, p)$. Here we estimate the number of edges in each part $U \cap V_i$ and between two different parts respectively:

$$|e(U) - \binom{n/2}{2}p| \leq \sum_{i=1}^t |e(U \cap V_i) - \binom{|U \cap V_i|}{2}p| + \sum_{i,j} |e(U \cap V_i, U \cap V_j) - |U \cap V_i||U \cap V_j|p|,$$

which by the fact that $e(U \cap V_i, U \cap V_j) = |U \cap V_i||U \cap V_j|p + O(\varepsilon)$ for all but at most $O(\varepsilon)t^2$ pairs $i,j$, is at most

$$\sum_{i=1}^t |V_i|^2 + \left(\sum_{i,j} |U \cap V_i||U \cap V_j|\right)(|p' - p| + O(\varepsilon)) + O(\varepsilon n^2) \leq n^2/t + O(\varepsilon|U|^2/2) + O(\varepsilon n^2) = O(\varepsilon n^2).$$

In the last equation, we took $t$ to be sufficiently large depending on $\varepsilon$ and $p$. Therefore by the quasi-randomness of $\mathcal{P}_2(1/2)$, we can conclude that $G$ is a quasi-random graph.

4. GENERAL CLIQUES

Throughout the section, $k$ and $r$ are fixed integers satisfying $r \geq k \geq 4$. Let an $r$-balanced cut be a $(1/r, \ldots, 1/r)$-cut. In this section, we will prove the remaining cases of the main theorem, quasi-randomness of general $k$-clique $r$-balanced cut properties.

**Theorem 4.1.** Fix a real $p \in (0, 1)$ and positive integers $r, k$ such that $r \geq k \geq 4$. For every $\varepsilon > 0$, there exists a positive real $\delta$ such that the following is true. If $G$ is a graph which has density $p$ and is $\delta$-close to satisfying the $K_k$ balanced cut property $C_k(1/r, \ldots, 1/r)$, then $G$ is $\varepsilon$-close to being $p$-quasi-random.

Let $G$ be a graph which is $\delta$-close to satisfying the $k$-clique $r$-balanced cut property. Apply the regularity lemma (Theorem 2.1) to this graph to obtain an $\varepsilon$-regular partition $\{V_i\}_{i=1}^r$ of the vertex set. For $i \in [r]$, let $x_i$ be the density of the edges within $V_i$, and for distinct $i, j \in [r]$, let $d_{ij}$ be the density of the pair $(V_i, V_j)$. For $k \geq 2$, a $k$-tuple $J = \{i_1, \ldots, i_k\}$ is a multiset of $k$-indices (not necessarily distinct). Let $d_J$ be the density of $k$-cliques which have exactly one vertex in each of the $V_{ia}$ for $a = 1, \ldots, k$. A $k$-tuple $J$ is called regular if $(V_{ia}, V_{ib})$ forms an $\varepsilon$-regular pair for all $a, b \in [k]$. For a $k$-tuple $J$ and a cut $X = \{X_1, \ldots, X_r\}$, we say that $J$ crosses the cut $X$ if $|J \cap X_i| \leq 1$ for all $i$, and denote it by $J \cap X$.

The proof of the $k$-clique $r$-balanced cut case follows the same line of the proof of the triangle case. First we develop two lemmas which correspond to Lemma 3.3 and Lemma 3.4. The following lemma can be regarded as generalization of Lemma 3.3 for arbitrary integers $k \geq 3$.
Lemma 4.2. For any \( k \geq 3 \), let \( \varepsilon \) be small enough depending on the densities \( d_{ij} \) for all \( i,j \in [t] \). There exists a function \( f : \mathbb{R} \to \mathbb{R} \) such that the following holds. Let \( J \) be a regular \( k \)-tuple, \( J' \subset J \) be such that \( |J'| = k - 2 \), and \( \{j_1,j_2\} = J \setminus J' \). Then,

\[
 x_{j_1} \left( \prod_{a \in J'} d_{a j_1} \right)^2 \left( \prod_{a,b \in J', a < b} d_{ab} \right) + x_{j_2} \left( \prod_{a \in J'} d_{a j_2} \right)^2 \left( \prod_{a,b \in J', a < b} d_{ab} \right) - 2 \left( \prod_{a,b \in J, a < b} d_{ab} \right) = \varepsilon + \delta f(t) O(1). 
\]

Proof. For the sake of clarity, without loss of generality we consider the case \( \{1,2\} \subset J \) and \( j_1 = 1, j_2 = 2 \). As in the triangle case, by considering the family of \( r \)-partitions

\[
\pi_a = ((1 - \alpha)V_1 + \alpha V_2, \alpha V_1 + (1 - \alpha)V_2, V_3, \ldots, V_t),
\]

and the density vector \( (d'_j)_{J \in \binom{[k]}{3}} \) which arise from these partitions, we can define \( d' = (d'_j)_{J \in \binom{[k]}{3}} \) as \( d'_j = d_j^{1/2} - \frac{1}{2} d_0^{1/2} - \frac{1}{2} d_1^{1/2} \). The same proof as in Lemma 3.2 gives us,

\[
d_j' = \begin{cases} 
0 & \text{if } J \cap \{1,2\} = \emptyset, \\
\alpha(1) & \text{if } J \cap \{1,2\} = \{1\}, \\
\alpha(1) & \text{if } J \cap \{1,2\} = \{2\}, 
\end{cases} \tag{7}
\]

and if \( \{1,2\} \subset J \) (let \( J' = J \setminus \{1,2\} \)) and \( J \) is a regular \( k \)-tuple, we have

\[
d_j' = \alpha(1 - \alpha) \sum_{i=1}^2 x_i \left[ \prod_{a \in J'} d_{ai} \right]^2 \left( \prod_{a,b \in J', a < b} d_{ab} \right) - 2 \left( \prod_{a,b \in J, a < b} d_{ab} \right) + O(\varepsilon). \tag{8}
\]

If \( \{1,2\} \subset J \) and \( J \) is not a regular \( k \)-tuple, then we do not have any control on \( d_j' \).

Let \( M \) be the \( \binom{[t]}{r} \times \binom{[k]}{r} \) \( 0 \)-1 matrix whose rows are indexed by \( r \)-balanced cuts of the vertex set \([t]\) and columns are indexed by the \( k \)-tuples \( \binom{[k]}{r} \). The \( (X_1, X_2, \ldots, X_r, J) \)-entry of \( M \) is 1 if and only if \( J \cap \{X_1, X_2, \ldots, X_r\} \). We know that \( M d' = \delta k \cdot 0 \).

Consider the submatrix \( N \) of \( M \) formed by the rows of partitions which have 1 and 2 in different parts, and columns of \( k \)-tuples which include both 1 and 2. Let \( d'' \) be the projection of \( d' \) onto the coordinates corresponding to the \( k \)-tuples which contain both 1 and 2. By \( M d'' = \delta k \cdot 0 \), we can conclude that \( N d'' = \delta k \cdot 0 \) given that the graph is large enough, since by (7) only the columns \( J \) with \( \{1,2\} \subset J \) will play a role here. Thus if we can show that \( N \) has full rank, then this implies that \( d_j' = \delta f(t) O(1) \) for all \( \{1,2\} \subset J \) (appeal to property (6) of \( \varepsilon \)-equality - the entries of \( N \) are bounded and the size of \( N \) depends on \( t \)).

Observe that the matrix \( N \) has a lot of repeated rows (we obtain identical rows by swaping elements from the parts containing 1 and 2 with each other). However, these repetitions do not contribute to the rank and can be ignored. Therefore we can consider a \( 0 \)-1 matrix \( N' \) whose rows are indexed by the collection of subsets \( Y = \{Y_1, \ldots, Y_{r-2}\} \) of the set \( T = \{3, \ldots, t\} \) where each part has size \( t/r \), and columns are indexed by the \( (k - 2) \)-tuples \( J' \in \binom{T}{k-2} \), where the entry \( (Y, J') \) is 1 if and only if \( J' \cap Y \). By fixing a subset of \( T \) of size \( (r - 2)t/r \) and considering all possible \( Y \) arising within this set, one can see that the row-space of \( N' \) generates the row-space of the \( (r - 2)t/r \) versus \( k - 2 \) inclusion matrix of \( T \), which we know by Gottlieb’s theorem, Theorem 2.7, has full rank. This implies that \( N \) has full rank as well.
Even though the equation which we obtained in Lemma 4.2 looks a lot more complicated than the triangle case, as it turns out, it is possible to make a substitution of variables so that the equations above become exactly the same as the equations in the triangle case. For a regular \((k - 3)\)-tuple \(I\), an index \(j \notin I\), and distinct \(j_1, j_2 \notin I\), define

\[
d_{j_1j_2} := d_{j_1j_2} \left( \prod_{a \in I} d_{aj} \right)^{1/2} \left( \prod_{a \notin I} d_{aj} \right)^{1/2} \left( \prod_{a, b \in I, a < b} d_{ab} \right)^{1/3}
\]

and

\[
x_j := x_j \left( \prod_{a \in I} d_{aj} \right) \left( \prod_{a, b \notin I, a < b} d_{ab} \right)^{1/3}.
\]

**Claim 4.3.** Let \(J\) be a regular \(k\)-tuple, \(I \subset J\) be of size \(|I| = k - 3\), \(\{j_1, j_2, j_3\} = J \setminus I\), and \(J' = I \cup \{j_3\}\). Then

\[
d_{j_1j_2} d_{j_3j_1} d_{j_2j_3} = \prod_{a, b \in I, a < b} d_{ab}, \quad \text{and} \quad x_{j_1} (d_{j_1j_3})^2 = x_{j_1} \left( \prod_{a \in I} d_{aj} \right)^2 \left( \prod_{a, b \notin I, a < b} d_{ab} \right).
\]

**Proof:** The claim follows from a direct calculation.

In other words, Claim 4.3 transforms the computation of the density of \(K_5\) in the graph into the computation of the density of triangles in another graph. This observation will greatly simplify the equations obtained from Lemma 4.2.

**Lemma 4.4.** Let \(\varepsilon\) be small enough depending on the densities \(d_{ij}\) for all \(i, j \in [t]\). There exists a function \(f : \mathbb{R} \to \mathbb{R}\) such that the following holds. Let \(J\) be a regular \(k\)-tuple and \(I \subset J\) be of size \(|I| = k - 3\). For \(\{j_1, j_2, j_3\} = J \setminus I\), we have \(x_{j_1} (d_{j_1j_3})^2 + x_{j_2} (d_{j_2j_3})^2 - 2d_{j_1j_2} d_{j_2j_3} d_{j_3j_1} = \varepsilon + \delta f(\varepsilon) 0\).

**Proof:** This is an immediate corollary of Lemma 4.2 and Claim 4.3.

The next lemma corresponds to Lemma 3.4 and establishes a lower bound on the densities. We omit the proof which is a straightforward generalization of the proof of Lemma 3.4.

**Lemma 4.5.** There exists \(t_0\) such that for fixed \(p \in (0, 1)\) and every \(t \geq t_0\) divisible by \(2r\), there exist \(c = c(k, p)\) and \(\delta_0 = \delta_0(p, t) > 0\) so that the following holds for every \(\delta \leq \delta_0\). If \(G\) is a graph with density \(p\) which is \(\delta\)-close to satisfying the \(k\)-clique balanced cut property, then for any partition \(\pi\) of \(V(G)\) into \(t\) equal parts, the density vector \(d = (d_{ij})_{i,j}\) satisfies \(d_{ij} \geq c\) for all distinct \(i, j \in [t]\).

For every fixed regular \((k - 3)\)-tuple \(I\), the set of equations that Lemma 4.4 gives is exactly the same as the set of equations obtained from Lemma 3.3. Consequently, by using Lemma 4.5, we can solve these equations for every fixed \(I\) just as in the triangle case. Note that there is no need to (re)develop a statement corresponding to Lemma 3.5, since now that we reduced our problem to the triangle case, the same lemma can be used as it is stated.

Thus as promised, we can reduce the case of general cliques to the case of triangles. Therefore the proof of the triangle case of Theorem 3.1 can be repeated to give us useful
information. However, this observation does not immediately imply that $d_{ij} = \varepsilon_0 p$ for most of the pairs $j_1, j_2 \not\in I$, since the only straightforward conclusion that we can draw is that for every regular $(k - 3)$-tuple $I$, there exists a constant $p_I$ such that $d_{ij} = \varepsilon_0 p_I$ for most of the pairs $j_1, j_2 \not\in I$. In order to prove the quasi-randomness of balanced cut properties, we will need some control on the relation between different $p_I$. Call a $k$-tuple $J$ excellent if it is regular, and for every $(k - 3)$-tuple $I \subset J$, we have $d_{ij} = \varepsilon_0 p_I$ for all distinct $j_1, j_2 \in J \setminus I$.

**Lemma 4.6.** Let $J$ be an excellent $k$-tuple. Then the density of every two pairs in $J$ are $\varepsilon$-equal to each other.

**Proof.** For the sake of clarity, assume that $J = \{1, 2, \ldots, k\}$. First, consider $I = \{4, \ldots, k\}$. Then by the assumption, we have $d_{13} = d_{23}$, which by definition gives,

$$d_{13} \left( \prod_{a \in I} d_{ia} \right)^{1/2} \left( \prod_{a \in I} d_{ia} \right)^{1/2} \left( \prod_{a \in I, a < b} d_{ab} \right)^{1/3} = \varepsilon d_{23} \left( \prod_{a \in I} d_{ia} \right)^{1/2} \left( \prod_{a \in I} d_{ia} \right)^{1/2} \left( \prod_{a \in I, a < b} d_{ab} \right)^{1/3}. $$

After cancelation of the same terms, we can rewrite this as,

$$d_{13} \left( \prod_{a=4}^k d_{ia} \right)^{1/2} = \varepsilon d_{23} \left( \prod_{a=4}^k d_{ia} \right)^{1/2} \iff d_{13}^{1/2} \left( \prod_{a=4}^k d_{ia} \right)^{1/2} = \varepsilon d_{23}^{1/2} \left( \prod_{a=4}^k d_{ia} \right)^{1/2}. \quad (9)$$

We can replace 3 by $i$ for $i \in \{3, 4, \ldots, k\}$ and multiply each side of all these equations to obtain,

$$\prod_{i=3}^k \left( d_{1i}^{1/2} \left( \prod_{a=3}^k d_{ia} \right)^{1/2} \right) = \prod_{i=3}^k \left( d_{2i}^{1/2} \left( \prod_{a=3}^k d_{ia} \right)^{1/2} \right),$$

which is equivalent to

$$\left( \prod_{i=3}^k d_{1i} \right)^{(k-1)/2} = \varepsilon \left( \prod_{i=3}^k d_{2i} \right)^{(k-1)/2}. $$

If we plug this back into Eq. (9), we get $d_{13} = \varepsilon d_{23}$. By repeating this process for other choice of indices, we can conclude that the density of every two pairs are $\varepsilon$-equal to each other.

We now combine all these observations to show that $d_e = \varepsilon_0 p$ for most of the edges $e$ of the reduced graph, which will in turn imply the quasi-randomness.

**Proof of Theorem 4.1.** Choose $\varepsilon_0$ small enough depending on the constant $c = c_{4.5}(p)$ so that the condition of Lemma 4.4 holds, and let $f$ be the function from Lemma 4.4. Let $\varepsilon \leq \min\{\varepsilon_0, 1/4\}, t_0 = t_{4.5}$, and let $T = T_{2.1}(\varepsilon, t_0)$. Let $\delta = \min_{0 \leq t \leq T} \{\varepsilon / f(t), \delta_{4.5}(t, p)\}$.

Let $G$ be a graph which is $\delta$-close to satisfying the $k$-clique $r$-balanced cut property. Apply the regularity lemma (Theorem 2.1) to this graph to obtain an $\varepsilon$-regular partition $\{V_i\}_{i=1}^t$ of the vertex set where $t$ is divisible by $2r$. For distinct $i, j \in [t]$, let $d_{ij}$ be the density of the pair $(V_i, V_j)$. Note that the parameters are chosen so that Lemma 4.4 and Lemma 4.5 holds.

Random Structures and Algorithms DOI 10.1002/rsa
For every regular \((k - 3)\)-tuple \(I\), define a graph \(H_I\) as following. The vertex set of \(H_I\) is the collection of elements of \([t]\) \(\setminus I\) which form a regular \((k - 2)\)-tuple together with \(I\). And \(j_1, j_2 \in V(H_I)\) forms an edge if and only if the \((k - 1)\)-tuple \(I \cup \{j_1, j_2\}\) is regular. Since each part of the regular partition forms a regular pair with at least \((1 - \epsilon) r\) of the other parts, we know that the graph \(H_I\) has at least \((1 - k \epsilon) r\) vertices and minimum degree at least \((1 - 2k \epsilon) r\). Thus by Lemma 4.4, Lemma 4.5, Lemma 3.5 and the proof of Theorem 3.1, we know that there exists a \(p_I\) such that at least \((1 - O(\epsilon))\)-proportion of the edges of \(H_I\) have density \(\epsilon\)-equal to \(p_I\).

Select \(k\) indices \(j_1, \ldots, j_k\) out of \([t]\) independently and uniformly at random. With probability at least \(1 - O(\epsilon)\), the \(k\)-tuple is regular. Moreover, with probability at least \(1 - O(\epsilon)\),

\[
d_{ij_{j_1}} = \epsilon^2 p_{ij_{j_1}}
\]

and the same is true for other choices of indices as well. Therefore by the union bound, the \(k\)-tuple \(\{j_1, \ldots, j_k\}\) is excellent with probability at least \(1 - O(\epsilon)\). Equivalently, the number of excellent \(k\)-tuples is at least \(\left(1 - O(\epsilon)\right)\binom{t}{k}\).

Call a pair of indices in \([t]\) excellent if it is contained in at least \(\frac{2}{3}\binom{t}{k-2}\) excellent \(k\)-tuples. Assume that there are \(\eta t^2\) non-excellent edges. Then the number of non-excellent \(k\)-tuples are at least

\[
\eta t^2 \times \frac{1}{3} \binom{t}{k-2}/\binom{k}{2} = \Omega(\eta) \binom{t}{k}.
\]

Therefore, \(\eta = O(\epsilon)\) and there are at most \(O(\epsilon) t^2\) non-excellent edges. We claim that all the excellent edges are \(\epsilon\)-equal to each other. Take two excellent edges \(e, f\). Since each of these edges form an excellent \(k\)-tuple with more than \(\frac{2}{3}\binom{t}{k-2}\) of the \((k - 2)\)-tuples, there exists a \((k - 2)\)-tuple which forms an excellent \(k\)-tuple with both of these edges. Thus by Lemma 4.6 applied to each of these \(k\)-tuples separately, we can conclude that \(d_e = \epsilon d_f\).

Consequently, by the total transitivity of \(\epsilon\)-equality (see, Section 2), we can conclude that \(d_e = \epsilon p'\) for some \(p'\) for every excellent edge \(e\). Then apply the same reasoning as in the triangle case to show that \(p' = \epsilon p\) and \(G \in \mathcal{P}_2(1/2)\). This proves the quasi-randomness of the graph \(G\).

5. CONCLUDING REMARKS

In this paper, we proved the quasi-randomness of \(k\)-clique balanced cut properties for \(k \geq 3\) and thus answered an open problem raised by both Shapira-Yuster [15] and Janson [11]. The most important base case was \(k = 3\) where we solved a system of equations given by Lemma 3.3. The existence of “bad” vertex in Lemma 3.5 complicated the proof of the main theorem. It is hard to believe that the case can be significantly simplified since even if we assume that all the pairs are regular in the regular partition, there is an assignment of variables \(x_i\) and \(d_i\) which is not all constant but forms a solution of the system.

We conclude this paper with an open problem for further study. It is a generalization of balanced cut property to counting general graph \(H\), which can also be regarded as an analogue of Simonovits and Sós’ theorems for cuts.

**Question 5.1.** Let \(k, r\) be positive integers satisfying \(r \geq k \geq 3\). Let \(H\) be a nonempty graph on \(k\) vertices, and assume that every \((1/r, \ldots, 1/r)\)-cut of a graph \(G\) has the “correct” number of copies of \(H\) such that every vertex of \(H\) is in a different part of the cut. Does this condition force \(G\) to be quasi-random?

*Random Structures and Algorithms* DOI 10.1002/rsa
One might be able to adapt our approach to solve this question. The main obstacle in this approach lies in the fact that the new system of equations we get as in Lemma 3.3 now become much more complicated to control. In order to characterize the structure of densities, one will need to solve these system of equations and prove statements such as in Lemma 3.5 and Lemma 4.6.

ACKNOWLEDGMENTS

We would like to thank Asaf Shapira, and our advisor Benny Sudakov for the kindness and advice they provided that greatly helped us in doing this work. We also want to thank Svante Janson and the two anonymous referees for their valuable comments and corrections.

REFERENCES

[1] N. Alon and J. Spencer, The probabilistic method, 2nd edition, Wiley, New York, 2000.
[2] K. Azuma, Weighted sums of certain dependent random variables, Tôkuku Math J 19 (1967), 357–367.
[3] F. R. K. Chung, R. L. Graham, and R. M. Wilson, Quasi-random graphs, Combinatorica 9 (1989), 345–362.
[4] F. R. K. Chung, R. L. Graham, and R. M. Wilson, Quasi-random set systems, J Am Math Soc 4 (1991), 151–196.
[5] F. R. K. Chung and R. L. Graham, Quasi-random tournaments, J Graph Theory 15 (1991), 173–198.
[6] F. R. K. Chung and R. L. Graham, Quasi-random hypergraphs, Random Struct Algorithm 1 (1990), 105–124.
[7] F. R. K. Chung and R. L. Graham, Maximum cuts and quasi-random graphs, A. Frieze and T. Łuczak (Editors), Random Graphs, Vol. 2, Wiley, New York, 1992, pp. 23–33.
[8] D. H. Gottlieb, A class of incidence matrices, Proc Am Math Soc 17 (1966), 1233–1237.
[9] A. Hajnal and E. Szemerédi, Proof of a conjecture of Erdős, in Combinatorial Theory and Its Applications, P. Erdős, A. Rényi, and V. Sós (Editors), Vol. 2, Amsterdam, Netherlands: North-Holland, 1970, pp. 601–623.
[10] W. Hoeffding, Probability inequalities for sums of bounded random variables, J Am Stat Assoc 58 (1963), 13–30.
[11] S. Janson, Quasi-random graphs and graph limits, Eur J Combin 32 (2011), 1054–1083.
[12] J. Komlós and M. Simonovits, Szemerédi’s regularity lemma and its applications in graph theory, Combinatorics, Paul Erdős is eighty, Vol. 2, Bolyai Society Mathematical Studies, János Bolyai Mathematical Society, Budapest, 1996, pp. 295–352.
[13] M. Krivelevich and B. Sudakov, Pseudo-random graphs, in More Sets, Graphs and Numbers, E. Győri, G. O. H. Katona, and L. Lovász (Editors), Vol. 15, Bolyai society mathematical studies, Springer, Berlin, 2006, pp. 199–262.
[14] C. McDiarmid, Concentration. In Probabilistic methods for algorithmic discrete mathematics, M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, and B. Reed (Editors), Springer, Berlin, 1998, pp. 1–46.
[15] A. Shapira and R. Yuster, The quasi-randomness of hypergraph cut properties, Random Struct Algorithm. DOI: 10.1002/rsa.20364.
[16] M. Simonovits and V. T. Sós, Hereditarily extended properties, quasi-random graphs and not necessarily induced subgraphs, Combinatorica 17 (1997), 577–596.

[17] E. Szemerédi, Regular partitions of graphs, In Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976), Volume 260: Colloquium International CNRS, CNRS, Paris, 1978, pp. 399–401.

[18] A. Thomason, Pseudo-random graphs, Proceedings of the Random Graphs, Poznań 1985, M. Karoński, editors, Ann of Discrete Math 33, North Holland, 1987, pp. 307–331.

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