Bipartite Subgraphs and Quasi-randomness

Jozef Skokan[1,*], Lubos Thoma[2,†]

[1] Department of Mathematics, University of Illinois at Urbana-Champaign, Urbana, IL 61820, USA, e-mail: `jozef@member.ams.org`

[2] Department of Mathematics, University of Rhode Island, Kingston, RI 02881, USA, e-mail: `thoma@math.uri.edu`

Abstract. We say that a family of graphs $\mathcal{G} = \{G_n : n \geq 1\}$ is p-quasi-random, $0 < p < 1$, if it shares typical properties of the random graph $G(n, p)$; for a definition, see below. We denote by $\mathcal{Q}^{\mathrm{w}}(p)$ the class of all graphs H for which $e(G_n) \geq (1 + o(1))p\binom{n}{2}$ and the number of not necessarily induced labeled copies of H in G_n is at most $(1 + o(1))p^{e(H)}n^{v(H)}$ imply that $\mathcal{G}$ is p-quasi-random. In this note, we show that all complete bipartite graphs $K_{a,b}$, $a, b \geq 2$, belong to $\mathcal{Q}^{\mathrm{w}}(p)$ for all $0 < p < 1$.

1. Notation

We start with fixing notation. For positive integers k, n and a real number x, we set $[n] = \{1, \ldots, n\}$ and $(x)_k = x(x-1) \times \cdots \times (x-k+1)$.

Given a graph G with vertex set $V(G)$ and edge set $E(G)$, $v(G)$ stands for $|V(G)|$ and $e(G)$ for $|E(G)|$. Furthermore, for a subset X of $V(G)$, $G[X]$ denotes the subgraph induced by the vertices of X, and $e(X)$ denotes the number of edges of $G[X]$. Given a vertex $x \in V(G)$, $N_G(x)$ is the set of all vertices adjacent to x and, similarly, for a subset X of $V(G)$, $N_G(X)$ denotes the set of all vertices adjacent to every vertex in X. Clearly, $N_G(X) = \bigcap_{x \in X} N_G(x)$. We also put $\deg(x) = \deg_G(x) = \big|N_G(x)\big|$ and $\deg(X) = \deg_G(X) = \big|N_G(X)\big|$.

For a graph $G = G_n$ on n vertices, let $\lambda_1(G), \ldots, \lambda_n(G)$,

$$\lambda_1(G) \geq |\lambda_2(G)| \geq \cdots \geq |\lambda_n(G)|\ ,$$

be the eigenvalues of its adjacency matrix.

Given two graphs G and H, a labeled *induced* copy of H in G is an injection $\psi \colon V(H) \to V(G)$ such that $\{x, x'\} \in E(H)$ if and only if $\{\psi(x), \psi(x')\} \in E(G)$.

Send offprint requests to: Lubos Thoma

* The first author was partially supported by NSF grant INT-0072064.

† The second author was partially supported by NSF grants DMS-9970622, DMS-0301228 and INT-0072064.

A labeled *weak* (i.e. not necessarily induced) copy of H in G is an injection $\psi\colon V(H) \to V(G)$ such that if $\{x,x'\} \in E(H)$, then $\{\psi(x),\psi(x')\} \in E(G)$. Denote by $\binom{G}{H}$ the set of all labeled induced copies of H in G and by $\binom{G}{H}^{\mathrm{w}}$ the set of all weak labeled copies of H in G.

2. Introduction

The theory of quasi-random graphs deals with properties of graphs, which are equivalent in the sense that a graph satisfying one of the properties must possess them all.

The study of quasi-random graphs was initiated by A.G. Thomason, cf. [11, 12], and systematically studied by Chung, Graham, and Wilson [5]. Their results were later extended to the case of uniform hypergraphs of a constant density, see [3,2,4,6].

Chung, Graham, and Wilson [5] proved the following theorem.

Theorem 1. *Let $0 < p < 1$, $\mathcal{G} = \{G_n : n \in \mathbb{N}\}$ be a family of graphs, $t \geq 4$ be an even integer, $s \geq 4$ be an integer. The following properties $\boldsymbol{P_1}$–$\boldsymbol{P_5}$ are equivalent for the family $\mathcal{G}$:*

$\boldsymbol{P_1(s)}$*: For all graphs H_s on s vertices,*

$$\left|\binom{G_n}{H_s}\right| = (1+o(1))p^{e(H_s)}(1-p)^{\binom{s}{2}-e(H_s)}n^s\ .$$

$\boldsymbol{P_2(t)}$*: $e(G_n) \geq (1+o(1))p\binom{n}{2}$ and $\left|\binom{G_n}{C_t}^{\mathrm{w}}\right| \leq (1+o(1))p^t n^t$, where C_t denotes a t-cycle.*

$\boldsymbol{P_3}$*: $e(G_n) \geq (1+o(1))p\binom{n}{2}$ and $\lambda_1(G_n) = (1+o(1))pn, \lambda_i(G_n) = o(n)$ for $i \geq 2$.*

$\boldsymbol{P_4}$*: For all $X \subseteq V(G_n)$, $e(X) = \frac{p}{2}|X|^2 + o(n^2)$.*

$\boldsymbol{P_5}$*: For all but at most $o(n)$ vertices $x \in V(G_n)$, $\deg_{G_n}(x) = (1+o(1))pn$ and for all but at most $o(n^2)$ pairs of vertices $x,x' \in V(G_n)$, $\deg_{G_n}(x,x') = (1+o(1))p^2 n$.*

The equivalence of these properties is understood in the following sense. For two properties involving $o(1)$ terms $\boldsymbol{P} = \boldsymbol{P}(o(1))$ and $\boldsymbol{P'} = \boldsymbol{P'}(o(1))$, the implication "$\boldsymbol{P} \Rightarrow \boldsymbol{P'}$" means that for every $\varepsilon > 0$ there is a $\delta > 0$ so that any graph G_n satisfying $\boldsymbol{P}(\delta)$ must also satisfy $\boldsymbol{P'}(\varepsilon)$, provided $n > N_0(\varepsilon)$.

The families $\mathcal{G}$ satisfying properties $\boldsymbol{P_1} - \boldsymbol{P_5}$ above are called *p-quasi-random.* We also refer to any property equivalent to any of $\boldsymbol{P_1} - \boldsymbol{P_5}$ as a *p-quasi-random property.*

Since [5], many other p-quasi-random properties have been discovered (e.g. [9,10]). Given a graph $H = H_t$, let $\boldsymbol{P_w(H)}$ be the following property:

$$\boldsymbol{P_w(H)}:\ e(G_n) \geq (1+o(1))p\binom{n}{2} \text{ and } \left|\binom{G_n}{H}^{\mathrm{w}}\right| \leq (1+o(1))p^{e(H)}n^t.$$

One may ask to determine the class $\mathcal{Q}^{\mathrm{w}} = \mathcal{Q}^{\mathrm{w}}(p)$ of all graphs H for which $\boldsymbol{P_w(H)}$ is a p-quasi-random property. By the above theorem, all even cycles belong to this class. Chung, Graham, and Wilson further observed (cf. [5]) that $K_{2,t}$ belong to $\mathcal{Q}^{\mathrm{w}} = \mathcal{Q}^{\mathrm{w}}(p)$, however, odd cycles do not. As remarked in [1] one does not expect $\mathcal{Q}^{\mathrm{w}}$ to be a large class. In this note, we show that all complete bipartite graphs $K_{a,b}$, $a, b \geq 2$, belong to $\mathcal{Q}^{\mathrm{w}}(p)$ for all $0 < p < 1$ (cf. Theorem 2 below).

3. Bipartite Graphs

Our goal is to prove the following theorem. We will use Corollary 1 in its proof. The proof of the corollary is postponed to Section 4.

Theorem 2. *Let $a, b \geq 2$ be integers. For $0 < p < 1$, let $\mathcal{G}_p$ be a family of graphs $\{G_n \colon e(G_n) \geq (1+o(1))p\binom{n}{2}\}$. If for every $G_n \in \mathcal{G}_p$*

$$\left| \binom{G_n}{K_{a,b}}^{\mathrm{w}} \right| \leq (1+o(1))p^{ab}n^{a+b}, \tag{1}$$

then $\mathcal{G}_p$ is p-quasi-random.

Proof. Note that adding (or removing) $o(n^2)$ edges to $G_n \in \mathcal{G}$ does not affect whether $\mathcal{G}$ satisfies property $\boldsymbol{P_4}$ and, consequently, whether $\mathcal{G}$ is p-quasi-random or not. Further, it also doesn't change (1).

Hence, we may assume $\deg_{G_n}(x) > A_1$ for every vertex $x \in V(G_n)$, $\deg_{G_n}(x, x') > A_2$ for all pairs of vertices $x, x' \in V(G_n)$, and $\deg_{G_n}(X) > A_a$ for all subsets $X \in [V(G_n)]^a$. Here $[V(G_n)]^a$ stands for the set of all a-element subsets of $V(G_n)$ and A_1, A_2 and A_a are numbers given by Corollary 1. Indeed, let Y be any subset of $V(G_n)$ with $2(A_1+A_2+A_a)$ vertices and set $Z = V(G_n)\setminus Y$. We add to $E(G_n)$ any missing edge between Y and Z and any missing edge within Y. Altogether, we increase the number of edges by at most $2(A_1+A_2+A_a)\times n + 4(A_1+A_2+A_a)^2 = o(n^2)$. It is easy to see that every vertex (pair of vertices, or a-element subset, respectively) has more than A_1 (A_2, or A_a respectively) common neighbors in Y.

Then, note that

$$\left| \binom{G_n}{K_{a,b}}^{\mathrm{w}} \right| = \sum_{X \in [V(G_n)]^a} a!b! \binom{\deg_{G_n}(X)}{b}. \tag{2}$$

We use the convexity of $\binom{y}{b}$ [1] for $y \geq 0$, Jensen's inequality, and the fact that $\left|[V(G_n)]^a\right| = \binom{n}{a}$ to estimate the right-hand side of (2). Indeed,

$$\sum_{X \in [V(G_n)]^a} a!b! \binom{\deg_{G_n}(X)}{b} \geq a!b! \binom{n}{a} \binom{\sum_{X \in [V(G_n)]^a} \deg_{G_n}(X) / \binom{n}{a}}{b}. \tag{3}$$

[1] We say that function $\binom{y}{b} = (y)_b/b!$ is convex on interval $[0, \infty)$ if the function

$$g_b(y) = \begin{cases} 0 & \text{for } 0 \leq y \leq b-1, \\ (y)_b/b! & \text{for } y \geq b-1 \end{cases}$$

is convex (in the usual sense) for $y \geq 0$.

Note that, by double counting and the convexity of $\binom{y}{a}$,

$$\sum_{X\in[V(G_n)]^a} \deg_{G_n}(X) = \sum_{x\in V(G_n)} \binom{\deg_{G_n}(x)}{a} \geq n\binom{2e(G_n)/n}{a}.$$

Since $e(G_n) \geq (1+o(1))p\binom{n}{2}$, we obtain

$$\sum_{X\in[V(G_n)]^a} \deg_{G_n}(X) \geq (1+o(1))np^a\binom{n}{a}.$$

We combine this with (2) and (3) and conclude that

$$\left|\binom{G_n}{K_{a,b}}^{\mathrm{w}}\right| \geq (1+o(1))n^a\left((1+o(1))np^a\binom{n}{a}/\binom{n}{a}\right)_b = (1+o(1))p^{ab}n^{a+b}. \quad (4)$$

Comparing (4) with (1), we obtain an asymptotic equality in all inequalities.

Hence, by Corollary 1 (note that $\deg_{G_n}(x) > A_1$ for every vertex $x \in V(G_n)$ is necessary to verify assumption (i) of this proposition), we get that for all but $o(n)$ vertices $x \in V(G_n)$ we have $\deg_{G_n}(x) = (1+o(1))pn$. Similarly, for all but $o(n^a)$ sets $X \in [V(G_n)]^a$, we have $\deg_{G_n}(X) = (1+o(1))p^a n$.

Consequently,

$$\sum_{X\in[V(G_n)]^a} \binom{\deg_{G_n}(X)}{2} = (1+o(1))\binom{n}{a}p^{2a}\binom{n}{2}. \quad (5)$$

On the other hand, by a double counting,

$$\sum_{X\in[V(G_n)]^a} \binom{\deg_{G_n}(X)}{2} = \sum_{\{x,x'\}\in[V(G_n)]^2} \binom{\deg_{G_n}(x,x')}{a}. \quad (6)$$

The right-hand side of (6) can be estimated using Jensen's inequality again:

$$\sum_{\{x,x'\}\in[V(G_n)]^2} \binom{\deg_{G_n}(x,x')}{a} \geq \binom{n}{2}\binom{\sum_{\{x,x'\}\in[V(G_n)]^2}\deg_{G_n}(x,x')/\binom{n}{2}}{a}. \quad (7)$$

Since $\sum_{\{x,x'\}\in[V(G_n)]^2}\deg_{G_n}(x,x') = \sum_{y\in V(G_n)}\binom{\deg_{G_n}(y)}{2} = (1+o(1))p^2n\binom{n}{2}$, we obtain

$$\begin{aligned}\sum_{\{x,x'\}\in[V(G_n)]^2} \binom{\deg_{G_n}(x,x')}{a} &\geq \binom{n}{2}\binom{\sum_{\{x,x'\}\in[V(G_n)]^2}\deg_{G_n}(x,x')/\binom{n}{2}}{a} \\ &\geq \binom{n}{2}\binom{(1+o(1))p^2n\binom{n}{2}/\binom{n}{2}}{a} \qquad (8)\\ &= (1+o(1))\binom{n}{a}p^{2a}\binom{n}{2}.\end{aligned}$$

Comparing (5) and (8) yields an asymptotic equality in (7). Consequently, by Corollary 1 (note that $\deg_{G_n}(x,x') > A_2$ for every pair $x, x' \in V(G_n)$ is needed to verify assumption (i) of this corollary), we have $\deg_{G_n}(x, x') = (1 + o(1))p^2 n$ for all but at most $o(n^2)$ pairs x, $x' \in V(G_n)$. Then, $\mathcal{G}_p$ is quasi-random by property $\boldsymbol{P_5}$, Theorem 1. $\square$

We remark that the combined use of convexity and Corollary 1 was already considered in [8,7] to address quasi-randomness for sparse graphs.

4. A Variant of the Cauchy-Schwarz Inequality

Since the function $x^k, k \geq 2$, is a strictly convex function, we have

$$\frac{a_1^k + \cdots + a_n^k}{n} \geq \left(\frac{a_1 + \cdots + a_n}{n}\right)^k \tag{9}$$

by Jensen's inequality with equality iff $a_1 = a_2 = \cdots = a_n$. The proposition below shows that if equality in (9) holds asymptotically, then almost all a_i's are roughly equal to their average.

Proposition 1. *For every $\delta > 0$ and a positive integer $k \geq 2$ there exists $\varepsilon > 0$ such that for non-negative reals $a_1, a_2, \ldots, a_n$ satisfying*

(i) $\sum_{i=1}^n a_i \geq (1 - \varepsilon)na$, *and*
(ii) $\sum_{i=1}^n a_i^k < (1 + \varepsilon)na^k$,

we have $\left|\{i \ : \ |a - a_i| < \delta a\}\right| > (1 - \delta)n$.

Proof. We distinguish two cases: $k = 2$ and $k > 2$.

The first case ($k = 2$) was already considered by Kohayakawa, Rödl, and Sissokho in [7]. We include their proof for the sake of completeness. Given $\delta > 0$ and non-negative reals $a_1, a_2, \ldots, a_n$, we set $\varepsilon = \delta^3/3$ and $B = \left|\{i \ : \ |a - a_i| \geq \delta a\}\right|$. We prove the proposition by showing $|B| < \delta n$.

Indeed, it follows from the definition of B that

$$\sum_{i=1}^n (a_i - a)^2 > |B|\delta^2 a^2. \tag{10}$$

By our assumption,

$$\begin{aligned}\sum_{i=1}^n (a_i - a)^2 &= \sum_{i=1}^n a_i^2 - 2a\sum_{i=1}^n a_i + \sum_{i=1}^n a^2 \\ &\leq (1+\varepsilon)a^2 - 2a(1-\varepsilon)na + na^2 = 3\varepsilon na^2. \end{aligned} \tag{11}$$

Combining (10) and (11) yields $|B|\delta^2 a^2 < 3\varepsilon na^2$, which implies $|B| < (3\varepsilon/\delta^2)n = \delta n$.

For the case when $k > 2$, we will use the well-known fact that for $k > 2$, we have

$$\left(\sum_{i=1}^n a_i^2/n\right)^{1/2} \leq \left(\sum_{i=1}^n a_i^k/n\right)^{1/k}. \tag{12}$$

Set $\varepsilon = \delta^3/3$. Then

$$\sum_{i=1}^{n} a_i^2 \overset{(12)}{\leq} \left(\sum_{i=1}^{n} a_i^k/n\right)^{2/k} n \overset{(ii)}{\leq} \left(\big(1+\delta^3/3\big)a^k\right)^{2/k} n \leq \big(1+\delta^3/3\big)a^2 n$$

because $2/k < 1$. Thus the proof follows from the $k = 2$ case. $\square$

The following corollary is similar to Proposition 1 but we consider function $\binom{x}{k}$ instead of x^k.

Corollary 1. *For every $0 < \delta < 1$ and positive integer $k \geq 2$ there exists $\varepsilon > 0$ and $A_k > 0$ such that for non-negative reals $a_1, a_2, \ldots, a_n, a$ satisfying*

(i) $a > A_k$ and $a_i > A_k$ for all $i \in [n]$,
(ii) $\sum_{i=1}^{n} a_i \geq (1-\varepsilon)na$, and
(iii) $\sum_{i=1}^{n} \binom{a_i}{k} < (1+\varepsilon)n\binom{a}{k}$,

we have $\big|\{i \ : \ |a - a_i| < \delta a\}\big| > (1-\delta)n$.

Proof. Set $\varepsilon = \delta^3/12$. Our conclusion follows from the fact that $a_1, \ldots, a_n$ and a satisfy the assumptions of Proposition 1. Since (ii) holds, we must only show that $\sum_{i=1}^{n} a_i^k \leq \big(1+\delta^3/3\big)na^k$.

Since $\lim_{x\to\infty} x^k/(x)_k = 1$ for every positive integer k, there exists a real number A_k such that $x^k < (1+\varepsilon)(x)_k$ whenever $x > A_k$. By (i)-(iii), we obtain

$$\sum_{i=1}^{n} a_i^k \overset{(i)}{<} (1+\varepsilon)\sum_{i=1}^{n}(a_i)_k = (1+\varepsilon)k!\sum_{i=1}^{n}\binom{a_i}{k} \overset{(iii)}{<} (1+\varepsilon)^2 k! n\binom{a}{k} \leq \big(1+\delta^3/3\big)na^k.$$

$\square$

5. Concluding remarks

In [5], the authors introduced forcing families of graphs. Let $\mathcal{G} = \{G_n : n \geq 1\}$ be a family of graphs. A family $\mathcal{F}$ of graphs is *p-forcing* if $\left|\binom{G_n}{F}^{\mathrm{w}}\right| = (1 + o(1))p^{e(F)}n^{v(F)}$ for all $F \in \mathcal{F}$ and $G_n \in \mathcal{G}$ implies $\mathcal{G}$ is p-quasi-random. Chung, Graham, and Wilson [5] asked what families were p-forcing, and, as an example of p-forcing families, they mentioned $\{P_2, C_{2t}\}$, $t \geq 2$, and $\{P_2, K_{2,t}\}$, $t \geq 2$.

Clearly, if H is any graph for which $\boldsymbol{P_w}(H)$ is p-quasi-random, then $\{P_2, H\}$ is p-forcing. In particular, $\{P_2, K_{a,b}\}$ is a p-forcing family for every a, $b \geq 2$. It would be interesting to decide whether $\{P_2, H\}$ is p-forcing (or $\boldsymbol{P_w}(H)$ is p-quasi-random) for every connected bipartite graph H with at least one cycle. We are not aware of an example of any bipartite graph H with at least one cycle for which $\boldsymbol{P_w}(H)$ is not p-quasi-random.

We also remark that if we consider induced copies instead of weak ones, then we do not obtain p-quasi-random properties. In particular, define the following property:

$$\boldsymbol{P}_{\mathbf{ind}}(H) : e(G_n) \geq (1+o(1))p\binom{n}{2} \quad \text{and}$$
$$\left|\binom{G_n}{H}\right| \leq (1+o(1))p^{e(H)}(1-p)^{\binom{v(H)}{2}-e(H)}n^{v(H)}.$$

Then, for every connected graph H, one can find a non-degenerate interval $I \subset [0,1]$ such that for each $p \in I$ there exists a family $\mathcal{G}_p = \{G_n : e(G_n) \geq (1+o(1))p\binom{n}{2}\}$ satisfying $\boldsymbol{P}_{\mathbf{ind}}(H)$ but $\mathcal{G}_p$ is not p-quasi-random. We remark that one can choose all $G_n \in \mathcal{G}_p$ of the form $G_n = 2G(n/2, q)$, where $G(n/2, q)$ is a random graph with edge probability q, and leave the details to the interested reader.

6. Acknowledgment

We would like to thank Andrew Thomason for fruitful discussions and Yoshi Kohayakawa for organizing Extended Workshop on Combinatorics in São Paulo, Ubatuba, and Rio de Janeiro, where a part of this work was done. We also thank the referees for their careful work.

References

1. B. Bollobás, *Random graphs* (2nd edition), Cambridge Studies in Advanced Mathematics 73, Cambridge Univ. Press, xviii+498pp., 2001.
2. F. R. K. Chung, *Quasi-random classes of hypergraphs*, Random Structures and Algorithms **1** (1990), no. 4, 363–382.
3. F. R. K. Chung and R. L. Graham, *Quasi-random hypergraphs*, Random Structures and Algorithms **1** (1990), no. 1, 105–124.
4. ———, *Quasi-random set systems*, Journal of the American Mathematical Society **4** (1991), no. 1, 151–196.
5. F. R. K. Chung, R. L. Graham, R. M. Wilson, *Quasi-random graphs*, Combinatorica **9** (1989), no. 4, 345–362.
6. Y. Kohayakawa, V. Rödl, and J. Skokan, *Hypergraphs, quasi-randomness, and conditions for regularity*, J. Combin. Theory Ser. A. **97** (2002), no. 2, 307–352.
7. Y. Kohayakawa, V. Rödl, and P. Sissokho, *Embedding graphs with bounded degree in pseudorandom sparse graphs*, submitted, 2002.
8. T. Łuczak, V. Rödl, and P. Sissokho, unpublished manuscript, 2000.
9. M. Simonovits and V. Sós, *Szemerédi's partition and quasi-randomness*, Random Structures and Algorithms, **2** (1991), no. 1, 1–10.
10. ———, *Hereditarily extended properties, quasi-random graphs and not necessarily induced subgraphs*, Combinatorica, **17** (1997), no. 4, 577-596.
11. A. G. Thomason, *Pseudorandom graphs*, Random graphs '85 (Poznań, 1985), North-Holland Math. Stud., vol. 144, North-Holland, Amsterdam–New York, 1987, pp. 307–331.

12. ______, *Random graphs, strongly regular graphs and pseudorandom graphs*, Surveys in Combinatorics 1987 (C. Whitehead, ed.), London Mathematical Society Lecture Note Series, vol. 123, Cambridge University Press, Cambridge–New York, 1987, pp. 173–195.