A Generalization of Turán's Theorem

Benny Sudakov,[1] Tibor Szabó,[2] and Van H. Vu[3]

[1]DEPARTMENT OF MATHEMATICS
PRINCETON UNIVERSITY
PRINCETON, NEW JERSEY
E-mail: bsudakov@math.princeton.edu

[2]DEPARTMENT OF COMPUTER SCIENCE
ETH ZÜRICH, 8092 SWITZERLAND
E-mail: szabo@inf.ethz.ch

[3]DEPARTMENT OF MATHEMATICS
UCSD, LA JOLLA, CALIFORNIA 92093
E-mail: vanvu@ucsd.edu

Received December 2, 2002; Revised September 7, 2004

Published online 29 March 2005 in Wiley InterScience(www.interscience.wiley.com).
DOI 10.1002/jgt.20074

Abstract: In this paper, we obtain an asymptotic generalization of Turán's theorem. We prove that if all the non-trivial eigenvalues of a d-regular graph G on n vertices are sufficiently small, then the largest K_t-free subgraph of G contains approximately $(t-2)/(t-1)$-fraction of its edges. Turán's theorem corresponds to the case $d = n - 1$. J Graph Theory 49: 187–195, 2005

Contract grant sponsor: NSF; Contract grant numbers: DMS-0355497, DMS-0106589 (to B.S.), DMS-0200357 (to V.H.V.); Contract grant sponsor: an Alfred P. Sloan Fellowship (to B.S. and V.H.V.); Contract grant sponsor: Joint Berlin/Zurich Graduate Combinatorics, Geometry, Computation (financed by the German Science Foundation (DFG) and ETH Zürich (to T.S.); Contract grant sponsor: UCSD; Contract grant number: RB091G-VU (to V.H.V.).

Keywords: *extremal graph theory; Turán's theorem; spectra of graphs*

1. INTRODUCTION

Turán's theorem [10] is one of the fundamental results in Extremal Graph Theory. It states that among n-vertex graphs not containing a clique of size t, the complete $(t-1)$-partite graph with (almost) equal parts has the maximum number of edges. For two graphs G and H, we define the Turán number $ex(G,H)$ of H in G, as the largest integer e, such that there is an H-free subgraph of G with e edges. Obviously, $ex(G,H) \leq |E(G)|$, where $E(G)$ denotes the edge set of G. Turán's theorem, in an asymptotic form, can be restated as

$$ex(K_n, K_t) = \left(\frac{t-2}{t-1} + o(1)\right)\binom{n}{2}, \tag{1}$$

that is, the largest K_t-free subgraph of K_n contains approximately $(t-2)/(t-1)$-fraction of its edges. We would like to extend this result to graphs other than K_n.

Let us consider an arbitrary graph G on n vertices. It is easy to give a lower bound on $ex(G, K_t)$ following Turán's construction. One can partition the vertex set of G into $t-1$ parts such that the degree of each vertex within its own part is at most $1/(t-1)$-times its degree in G. Thus, the subgraph consisting of the edges of G connecting two different parts has at least a $(t-2)/(t-1)$-fraction of the edges of G and is clearly K_t-free. We say that a graph (or rather a family of graphs) is *t-Turán* if this trivial lower bound is essentially an upper bound as well. More precisely, G is t-Turán if $ex(G,K_t) = \left(\frac{t-2}{t-1} + o(1)\right)|E(G)|$. The question we pursue is

Which graphs are t-Turán ?

It has been shown that for any fixed t, there is a number $m(t,n)$ such that almost all graphs on n vertices with $m \geq m(t,n)$ edges are t-Turán. The most recent estimate for $m(t,n)$, due to Szabó and Vu [9], is $cn^{2-\frac{1}{t-1.5}}$, provided $t \geq 4$ and c is a sufficiently large constant. It is conjectured that one can set $m(t,n)$ as small as $c\,n^{2-2/(t+1)}$, but so far this has been verified only for $t = 3$ [4], $t = 4$ [5], and $t = 5$ [6]. All these results, however, are results about random graphs and do not yet provide a deterministic sufficient condition for a graph to be t-Turán.

The main difficulty in generalizing Turán's theorem is that all of its classical proofs are tailored to the complete graph K_n. However, the recent investigations in [9] revealed that one of the key conditions for a graph to be t-Turán is that its edges are distributed sufficiently evenly. It has turned out that under certain circumstances, this condition (to be more precise, a sufficiently strong variant of it) can be guaranteed by a simple assumption about the spectrum of the graph.

For a graph G, let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be the eigenvalues of its adjacency matrix. The quantity $\lambda(G) = \max\{\lambda_2, -\lambda_n\}$ is called the *second eigenvalue* of G. A graph $G = (V,E)$ is called an (n,d,λ)*-graph* if it is d-regular and has n

vertices, and the second eigenvalue of G is at most λ. It is well known (see [3] for more details) that if λ is much smaller than the degree d, then G has certain random-like properties. Thus, λ could serve as some kind of "measure of randomness" in G. Our main result is the following:

Theorem 1.1. *Let $t \geq 3$ be an integer and let $G = (V, E)$ be an (n, d, λ)-graph. If $d^{t-1}/n^{t-2} \gg \lambda$, then*

$$ex(G, K_t) = \left(\frac{t-2}{t-1} + o(1)\right)|E(G)|.$$

To see that this result generalizes Turán's theorem, observe that the second eigenvalue of the complete graph K_n is 1. Theorem 1.1 could also be considered as a contribution to the fast-developing comprehensive study of graph theoretical properties of (n, d, λ)-graphs, which has recently attracted lots of attention both in combinatorics and theoretical computer science. For a recent survey about these fascinating graphs and their properties, we refer the interested reader to the paper of Krivelevich and Sudakov [7].

Our proof of Theorem 1.1 uses an approach similar to that in [9] with some additional ideas. The main obstacle of an adaptation is that it is not clear how the technical condition used there can be applied for (n, d, λ)-graphs. The crucial Definition 2.3 of our paper circumvents this difficulty with only a slight loss in the outcome, and then a simplified variant of the double-counting argument of [9] can be applied.

Let us briefly discuss the sharpness of Theorem 1.1. The condition involving n, d, and λ is known to be tight (up to a constant factor) for $t = 3$. For $t = 3$, the above theorem states that if $d^2/n \gg \lambda$, then one needs to delete at least half of the edges of G to destroy all the triangles. On the other hand, in [1], Alon constructed a *triangle-free* d-regular graph G on n vertices with second eigenvalue $\lambda = \lambda(G)$, where $d = \Theta(n^{2/3})$ and $\lambda = \Theta(n^{1/3})$. Using the blow-up of this construction, it was shown in [8] that for any pair of integers d and n such that $\Omega(n^{2/3}) \leq d \leq n$, there exist a triangle-free graph G_1, which has $n_1 = O(n)$ vertices, is d_1-regular with $d_1 = \Theta(d)$, and whose second eigenvalue satisfies $\lambda(G_1) = O(d_1^2/n_1)$. Obviously, $ex(G_1, K_3) = |E(G_1)|$. This implies that for $t = 3$ and any sensible degree d, the condition in Theorem 1.1 is not far from being best possible.

The rest of this paper is organized as follows. In the next section, we summarize some useful quantitative results on the edge distribution of pseudo-random regular graphs, which we use later in the proof. In Section 3, we present the proof of our main theorem. The last section of the paper is devoted to concluding remarks and discussion of relevant open problems.

We close this section with some conventions and notation. For two (not necessarily) disjoint subsets of vertices $U, W \subset V$, let $e(U, W)$ be the number of ordered pairs (u, w) such that $u \in U,\ w \in W$, and (u, w) is an edge of G. We also denote by $e(U) = e(U, U)/2$, the number of edges spanned by U. For a vertex v

of G, let $N(v)$ denote the set of vertices of G adjacent to v, and let $d(v)$ denote its degree. Similarly, for a subset U of the vertex set, $N_U(v) = N(v) \cap U$ and $d_U(v) = |N_U(v)|$. We will make no serious attempt to optimize our absolute constants.

2. PROPERTIES OF PSEUDO-RANDOM GRAPHS

In this section, we obtain a result on the edge distribution in pseudo-random regular graphs, which will be used later in the proof. We start with the following two well-known facts whose proofs can be found, e.g., in Chapter 9 of the monograph of Alon and Spencer [3].

Theorem 2.1. *Let $G = (V, E)$ be an (n, d, λ)-graph. Then for every subset U of V*

$$\sum_{v \in V} \Big(d_U(v) - d|U|/n\Big)^2 \leq \lambda^2 |U|.$$

This theorem has the following easy corollary (see, e.g., [3]).

Corollary 2.2. *Let $G = (V, E)$ be an (n, d, λ)-graph. Then for every two subsets $B, C \subset V$, we have*

$$\left| e(B, C) - \frac{|B||C|d}{n} \right| \leq \lambda \sqrt{|B||C|}.$$

To make our inductive argument work, we need the following somewhat technical definition.

Definition 2.3. *Let $G = (V, E)$ be a graph of order n, let $t \geq 2$ be an integer, and let $\delta(n)$ and $p(n)$ be two functions of n such that $0 < p = p(n) \leq 1$ and $\delta(n)$ tends to zero when n tends to infinity. We say that G has the (t, p, δ)-property if it satisfies the following two conditions*:

(*i*) *For every two subsets U and W of $V(G)$ of cardinality, at least $(\delta p)^{t-2} n$*

$$\Big| e(U, W) - p|U||W| \Big| \leq \delta p |U||W|.$$

(*ii*) *For every subset U of $V(G)$ with cardinality at least $(\delta p)^{t-3} n$, there are at most $(\delta p)^{t-1} n$ vertices of G with*

$$\Big| d_U(v) - p|U| \Big| > \delta p |U|.$$

The main result of this section provides a sufficient condition for an (n, d, λ)-graph to have the $(t, d/n, \delta)$-property.

Proposition 2.4. *Let $t \geq 2$ be an integer, and let $G = (V, E)$ be an (n, d, λ)-graph such that*

$$\frac{d^{t-1}}{n^{t-2}} > \omega(n)\lambda,$$

where $\omega(n)$ is a function, which tends to infinity arbitrarily slowly with n. Set $\delta = \delta(n) = \omega(n)^{-1/(t-1)}$. If n is sufficiently large, then G has the $(t, d/n, \delta)$-property.

Proof. (i) Let U and W be subsets of $V(G)$ with cardinality at least $(\delta d/n)^{t-2}n$. Then by Corollary 2.2, we have

$$\left| e(U,W) - \frac{d}{n}|U||W| \right| \leq \lambda\sqrt{|U||W|} = \frac{\lambda}{\sqrt{|U||W|}}|U||W| \leq \frac{\lambda}{(\delta d/n)^{t-2}n}|U||W|$$
$$= \frac{\lambda n^{t-2}}{\delta^{t-2}d^{t-1}}\frac{d}{n}|U||W| < \frac{1}{\omega(n)\delta^{t-2}}\frac{d}{n}|U||W| = \delta\frac{d}{n}|U||W|.$$

(ii) Let $U \subseteq V(G)$ with $|U| \geq (\delta d/n)^{t-3}n$. By Theorem 2.1, the number of vertices with $\left|d_U(v) - d|U|/n\right| > \delta d|U|/n$ is at most

$$\frac{\lambda^2|U|}{(\delta d|U|/n)^2} = \frac{\lambda^2 n^2}{\delta^2 d^2|U|} \leq \frac{\lambda^2 n^2}{\delta^2 d^2 (\delta d/n)^{t-3}n} = \frac{\lambda^2 n^{t-2}}{\delta^{t-1}d^{t-1}} = \frac{\lambda\,\omega(n)\lambda n^{t-2}}{d^{t-1}}$$
$$\leq \lambda \leq \frac{d^{t-1}}{\omega(n)n^{t-2}} = \left(\delta\frac{d}{n}\right)^{t-1}n.$$

■

3. PROOF OF THE MAIN RESULT

In this section, we prove that in graphs having the (t, p, δ)-property, Turán's theorem is valid asymptotically. In the light of Proposition 2.4, the following result (which may be of independent interest) immediately implies the assertion of Theorem 1.1.

Theorem 3.1. *Let $t \geq 2$ be an integer, and let $G = (V, E)$ be a graph of order n. If G has the (t, p, δ)-property with some $0 < p = p(n) \leq 1$ and $\delta = \delta(n) \to 0$, then*

$$ex(G, K_t) = \left(1 - \frac{1}{t-1} + o(1)\right)|E(G)|.$$

Proof. As it was noted in the Introduction, the lower bound $ex(G, K_t) \geq \frac{t-2}{t-1}|E(G)|$ is valid for every graph G.

To prove the corresponding upper bound on $ex(G, K_t)$, it is enough to show that for n large enough one needs to remove at least $\left(\frac{1}{t-1} - 9t\delta\right)n^2p/2$ edges from G in order to destroy all copies of K_t. Since by part (i) of Definition 2.3, G has at most $(1+\delta)n^2p/2$ edges, this indeed would imply that

$$\begin{aligned} ex(G, K_t) &\leq e(G) - \left(\frac{1}{t-1} - 9t\delta\right)\frac{n^2}{2}p \leq e(G) - \left(\frac{1}{t-1} - 9t\delta\right)\frac{e(G)}{1+\delta} \\ &= \left(\frac{t-2}{t-1} + o(1)\right)e(G). \end{aligned}$$

We prove the above claim by induction on t. For $t = 2$, this statement follows easily from part (i) of the definition of $(2, p, \delta)$-property. Indeed, G has at least $(1-\delta)n^2p/2$ edges, and we need to delete all of them to obtain a K_2-free graph.

Now let us assume that the claim holds for some $t \geq 2$ and prove it for $t+1$. Consider a graph G, which has the $(t+1, p, \delta)$-property, and let R be the set of edges of G such that deleting R destroys all $(t+1)$-cliques in G. Color the edges in R red and all other edges blue. Let N_1, N_2 be the number of triangles in G with exactly one and two red edges, respectively. To prove the claim, we will estimate N_1 from both sides.

For each vertex v, let R_v (B_v) be the set of neighbors of v, which are connected to v by a red (blue) edge and denote by $r_v = |R_v|$ $(b_v = |B_v|)$, the red (blue) degree of v. Let $r = \frac{1}{n}\sum_v r_v$ be the average red degree and $b = \frac{1}{n}\sum_v b_v$ be the average blue degree. Observe that $(1-\delta)np \leq r + b \leq (1+\delta)np$. Furthermore, let f_v be the number of red edges in the graph induced by B_v. It is clear that

$$N_1 = \sum_{v \in V} f_v\,. \tag{2}$$

For every vertex v of G, we distinguish two cases. First, assume $|B_v| \geq \delta np$. Then, it is easy to check that the induced subgraph $G[B_v]$ has the (t, p, δ)-property. As $E(G) - R$ does not contain K_{t+1}, the deletion of red edges should destroy all copies of K_t in $G[B_v]$. Therefore, by the induction hypothesis,

$$f_v \geq \left(\frac{1}{t-1} - 9t\delta\right)\frac{|B_v|^2}{2}p\,. \tag{3}$$

In the second case, when $|B_v| < \delta np$, we clearly have that

$$\left(\frac{1}{t-1} - 9t\delta\right)\frac{|B_v|^2}{2}p \leq \frac{\delta^2}{2}n^2p^3\,.$$

This, together with (2) and (3) implies the following lower bound on N_1

$$\begin{aligned}
2N_1 &\geq \sum_{v\in V}\left(\left(\frac{1}{t-1} - 9t\delta\right)b_v^2 p - \delta^2 n^2 p^3\right) \\
&= \left(\frac{1}{t-1} - 9t\delta\right)p\sum_{v\in V} b_v^2 - \delta^2 n^3 p^3 \\
&\geq \left(\frac{1}{t-1} - 9t\delta\right)b^2 np - \delta^2 n^3 p^3 \\
&\geq \frac{npb^2}{t-1} - \left(9t\delta(1+\delta)^2 + \delta^2\right)n^3 p^3. \qquad (4)
\end{aligned}$$

Here we used that by the Cauchy–Schwartz inequality $\sum_v b_v^2 \geq nb^2$ and that $b \leq (1+\delta)np$.

Now we obtain an upper bound on N_1.

$$2N_1 \leq 2N_1 + 2N_2 = \sum_{v\in V(G)} e(R_v, B_v). \qquad (5)$$

To estimate $e(R_v, B_v)$, we distinguish three cases. If b_v and r_v are both at least δpn, then by part (i) of the definition of the $(t+1, p, \delta)$-property, we have

$$e(R_v, B_v) \leq (1+\delta)pr_v b_v.$$

Now suppose that one of r_v and b_v, say r_v, is less than δnp. Assume, moreover, that $|d(v) - pn| \leq \delta pn$. Then we know that $d(v) = |R_v \cup B_v| \geq |B_v| \geq (1-2\delta)np \geq \delta np$ and so by part (i) of Definition 2.3,

$$\begin{aligned}
e(R_v, B_v) &= e(R_v \cup B_v) - e(R_v) - e(B_v) \\
&\leq (1+\delta)\frac{(r_v + b_v)^2 p}{2} - (1-\delta)\frac{b_v^2 p}{2} \\
&= (1+\delta)r_v b_v p + \delta p b_v^2 + (1+\delta)\frac{r_v^2 p}{2} \\
&\leq (1+\delta)r_v b_v p + 2\delta n^2 p^3.
\end{aligned}$$

Finally, assume that $r_v < \delta np$ and $|d(v) - pn| > \delta pn$. By part (ii) of the definition of the $(t+1, p, \delta)$-property, there are at most $(\delta p)^t n \leq \delta^2 p^2 n$ such vertices. For these v, $e(R_v, B_v) \leq r_v b_v \leq \delta pn \cdot n$, so altogether they contribute at most $\delta^3 p^3 n^3$ to the sum (5).

Summing up over all vertices the above estimates on $e(R_v, B_v)$, from (5) we obtain that

$$\begin{aligned}
2N_1 &\le \sum_{v\in V(G)} \Big((1+\delta)pr_v b_v + 2\delta n^2p^3\Big) + \delta^3p^3n^3 \\
&= (1+\delta)p \sum_{v\in V(G)} \Big(d(v)b_v - b_v^2\Big) + (2\delta+\delta^3)n^3p^3 \\
&\le (1+\delta)p \sum_{v\in V(G)} d(v)b_v - (1+\delta)b^2np + (2\delta+\delta^3)n^3p^3 \\
&\le (1+\delta)p \sum_{v\in V(G)} d(v)b_v - b^2np + 4\delta n^3p^3.
\end{aligned}$$

Here we used again that $\sum_v b_v^2 \ge nb^2,\ b \le (1+\delta)np$, and that $\delta^3 \ll \delta^2 \ll \delta$. Denote by V', the set of vertices with $|d(v)-pn| \le \delta pn$. Since $|V(G)\backslash V'| \le (\delta p)^t n \le \delta^2p^2n$, then

$$\begin{aligned}
2N_1 &\le (1+\delta)p\left(\sum_{v\in V'} d(v)b_v + \sum_{v\notin V'} d(v)b_v\right) - b^2np + 4\delta n^3p^3 \\
&\le (1+\delta)p\Big((1+\delta)np\cdot nb + \delta^2p^2n\cdot n^2\Big) - b^2np + 4\delta n^3p^3 \\
&= bn^2p^2 - b^2np + (2\delta+\delta^2)bn^2p^2 + (1+\delta)\delta^2n^3p^3 + 4\delta n^3p^3 \\
&\le npb(np-b) + 7\delta n^3p^3.
\end{aligned}$$

Combining this with (4), we have

$$\frac{npb^2}{t-1} - \Big(9t\delta(1+\delta)^2 + \delta^2\Big)n^3p^3 \le pbn(pn-b) + 7\delta n^3p^3. \tag{6}$$

We claim that $b \le \frac{t-1}{t}pn + (9t+8)\delta np$. Indeed, if $b \le \frac{t-1}{t}pn$, then we are done. Otherwise, dividing both parts of (6) by $npb \ge \frac{t-1}{t}n^2p^2$ and using that $\delta^3 \ll \delta^2 \ll \delta$, we obtain that

$$\frac{b}{t-1} \le (pn-b) + \frac{t}{t-1}(9t+8)\delta np.$$

Therefore

$$b \le \frac{t-1}{t}pn + (9t+8)\delta np.$$

Since $r+b \ge (1-\delta)np$, the last inequality implies that

$$|R| = \frac{nr}{2} \ge \frac{n\big((1-\delta)np-b\big)}{2} \ge \left(\frac{1}{t} - 9(t+1)\delta\right)\frac{n^2p}{2}.$$

This completes the proof of the induction step and the proof of the theorem. ∎

4. CONCLUDING REMARKS

The major question, of course, remains to determine a threshold-type condition for the validity of an asymptotic Turán's theorem in pseudo-random graphs. Theorem 1.1 together with Alon's construction [1] implies that $d^2n/\lambda = \Theta(1)$ is some kind of a threshold for an (n, d, λ) graph to be 3-Turán, but it is not clear what happens when $t \geq 4$. Some construction could be obtained by the slight modification of the Erdős-Rényi graphs, which appears in the paper of Alon and Krivelevich [2]. These are (n, d, λ)-graphs with parameters $d = n^{\frac{t-3}{t-2}}(1 + o(1))$ and $\lambda = n^{\frac{t-3}{2(t-2)}}(1 + o(1))$, which can be made K_t-free by deleting at most $n^{\frac{t-3}{t-2}}(1 + o(1))$ vertices. It shows that Theorem 1.1 is *not true* with the weaker condition $\lambda < Cd^{\frac{t-1}{2}}/n^{\frac{t-3}{2}}$, provided C is a large enough constant. A plausible approach to improve the condition of Theorem 1.1 could involve an adjusted double counting argument with a suitably chosen technical condition in the place of Definition 2.3. We believe, however, that such an improvement is not possible, and there is an extension of Alon's construction to K_t-free graphs. Note that for most values of t, such an extension would *not* have to improve on the known constructive bounds of the asymmetric Ramsey number $r(K_t, K_n)$.

REFERENCES

[1] N. Alon, Explicit Ramsey graphs and orthonormal labelings, Electron J Combin 1 (1994), R12.

[2] N. Alon and M. Krivelevich, Constructive bounds for a Ramsey-type problem, Graphs and Combinatorics 13 (1997), 217–225.

[3] N. Alon and J. H. Spencer, The Probabilistic Method, 2nd edition, Wiley, New York, 2000.

[4] P. Frankl and V. Rödl, Large triangle-free subgraphs in graphs without K_4, Graphs and Combinatorics 2 (1986), 135–144.

[5] Y. Kohayakawa, T. Łuczak, and V. Rödl, On K^4-free subgraphs of random graphs, Combinatorica 17 (1997), 173–213.

[6] S. Gerke, A. Steger, and T. Schickinger, K_5-free subgraphs of random graphs, Random Structures and Algorithms 24 (2004), 194–232.

[7] M. Krivelevich and B. Sudakov, Pseudo-random graphs, to appear.

[8] M. Krivelevich, B. Sudakov, and T. Szabó, Triangle factors in sparse pseudo-random graphs, Combinatorica 24 (2004), 403–426.

[9] T. Szabó and V. H. Vu, Turán's theorem in sparse random graphs, Random Structures and Algorithms 23 (2003), 225–234.

[10] P. Turán, Egy gráfelméleti szélsőértékfeladatról (in Hungarian), Mat Fiz Lapok 48 (1941), 436–452.