Threshold Functions for the Bipartite Turán Property

Anant P. Godbole

Department of Mathematical Sciences, Michigan Technological University, Houghton, MI 49931-1295, U.S.A. (anant@mtu.edu)

Ben Lamorte

Engineering and Economic Systems Department, Stanford University, Stanford, CA 93405-4025, U.S.A. (lamorte@leland.stanford.edu)

Erik Jonathan Sandquist

Department of Mathematics, Cornell University, Ithaca, NY 14853-7901, U.S.A. (ejs9@cornell.edu)

Submitted: May 20, 1997; Accepted: August 20, 1997

MR Subject Numbers: 05C50, 05C80, 05C35, 05B30

ABSTRACT

Let $G_2(n)$ denote a bipartite graph with n vertices in each color class, and let $z(n,t)$ be the bipartite Turán number, representing the maximum possible number of edges in $G_2(n)$ if it does not contain a copy of the complete bipartite subgraph $K(t,t)$. It is then clear that $\zeta(n,t)=n^2-z(n,t)$ denotes the minimum number of zeros in an $n\times n$ zero-one matrix that does not contain a $t\times t$ submatrix consisting of all ones. We are interested in the behaviour of $z(n,t)$ when both t and n go to infinity. The case $2\le t\ll n^{1/5}$ has been treated in [9] ; here we use a different method to consider the overlapping case $\log n\ll t\ll n^{1/3}$. Fill an $n\times n$ matrix randomly with z ones and $\zeta=n^2-z$ zeros. Then, we prove that the asymptotic probability that there are no $t\times t$ submatrices with all ones is zero or one, according as $z\ge(t/ne)^{2/t}\exp\{a_n/t^2\}$ or $z\le(t/ne)^{2/t}\exp\{(\log t-b_n)/t^2\}$, where a_n tends to infinity at a specified rate, and $b_n\to\infty$ is arbitrary. The proof employs the extended Janson exponential inequalities [1].

1. INTRODUCTION AND STATEMENT OF RESULTS

Given a graph F, what is the maximum number of edges in a graph on n vertices that does not contain F as a subgraph? In the bipartite case, we let $z(n,t)$ denote the (diagonal) bipartite Turán number, which represents the maximum number of edges in a bipartite graph [with n vertices in each color class] that does not contain a complete bipartite graph $K(t,t)$ of order t. An equivalent formulation of this problem is in terms of zero-one matrices, and is called the problem of Zarankiewicz: What is the smallest number of zeros $\zeta(n,t)$ that can be strategically placed among the entries of an $n \times n$ zero-one matrix so as to prevent the existence of a $t \times t$ submatrix of all ones? We remind the reader that, in this formulation, the submatrix in question need not have consecutive rows or columns. It is clear that $\zeta(n,t) = n^2 - z(n,t)$. [Generalizing this problem to $s \times t$ submatrices of a zero-one matrix of order $m \times n$ leads naturally to the numbers $z(m,n,s,t)$ and $\zeta(m,n,s,t)$; Bollobás [4]has shown that

$$2\mathrm{ex}(n, K(s,t)) \le z(n,n,s,t),$$

where $\mathrm{ex}(n,F)$ denotes the maximum number of edges in a graph on n vertices that does not contain F as a subgraph.] In contrast with the classical Turán numbers, definitive general results are not known in the bipartite case. The initial search for numerical values of $z(n,t)$, $t = 3,4,5\ldots; n = 4,5,6,\ldots$, due to Zarankiewicz; Sierpinski; Brzezinski; Čulik; Guy; and Znám, is chronicled in [4], as is the history of research (due to Hartman, Mycielski and Ryll-Nardzewski; and Rieman) leading to asymptotic bounds on $z(n,2)$, and on $z(m,n,s,t)$ (the latter set of results are due to Kövári, Sós and Turán; Hyltén-Cavallius; and Znám). The asymptotics of the numbers $z(n,n,2,t)$ (t fixed) and $z(n,3)$ have most recently been investigated by Füredi ([6], [7]) who also describes the early related work of Rieman; Kövári, Sós and Turán; Erdős, Rényi and Sós; Brown; Hyltén-Cavallius; and Mörs. An excellent survey of these and related questions can be found in Section VI.2 of [4]. A problem similar in spirit to the Zarankiewicz question is the object of intense study in reliability theory; see [2] for details and references, and [3] for background on the Stein-Chen method of Poisson approximation.

Most of the work described in the previous two paragraphs has focused on the case where the dimensions (s, t) of the forbidden submatrix are fixed, and n tends to infinity; a notable exception to this is provided by the recent work of Griggs and Ouyang [11] , and Gentry [8], who each study the half-half case, and derive several bounds and exact values for the numbers $z(2m, 2n, m, n)$. We continue this trend in this paper, focus on the diagonal case $m = n; s = t$, and study the asymptotics of the problem as *both* n and t tend to infinity. Our arguments will force us to assume that $\log n \ll t \ll n^{1/3}$, where, given two non-negative sequences a_n and b_n, we write $a_n \ll b_n$ if $a_n/b_n \to 0$ $(n \to \infty)$. We thus obtain an extension of the results in [9], where the overlapping case $2 \le t \ll n^{1/5}$ was considered. Similarities and differences between the approaches in [9] and the present paper will be given later in this section, and in the next section. Since $z(n, t) \sim n^2$ for the range of t's that we consider, we will occasionally rephrase our results in terms of the minimum number $\zeta(n, t)$ of zeros of an $n \times n$ 0-1 matrix that prevents the existence of a $t \times t$ submatrix of all ones. The key general bounds due to Znám [15] and Bollobás [4](Theorems VI.2.5 and VI.2.10 in [4], adapted to our purpose,) are as follows:

$$\left(n^2 - (t-1)^{1/t} n^{2-\frac{1}{t}} - \frac{n(t-1)}{2}\right) \le \zeta(n, t) \le \frac{2n^2 \log n}{t}\{1 + o(1)\} \ \ (t \to \infty; \ \ t \gg \log n), \tag{1}$$

In particular, with $t = n^\alpha, \alpha < 1/2$, we have

$$(1 - \alpha)n^{2-\alpha} \log n\{1 + o^*(1)\} \le \zeta(n, n^\alpha) \le 2n^{2-\alpha} \log n\{1 + o(1)\}. \tag{2}$$

We restate (1) and (2) in probabilistic terms as follows: Consider the probability measure $\mathbf{P}_{u,z}$ that randomly and uniformly places ζ zeros and $z = n^2 - \zeta$ ones among the entries of the $n \times n$ matrix [the subscript u refers to the fact that the allotment is uniform, and the subscript z to the fact that there are z ones in the array.] Let X denote the random variable that equals the number of $t \times t$ submatrices consisting of all ones [we often denote such a $t \times t$ matrix by J_t]. In other words,

$$X = \sum_{j=1}^{\binom{n}{t}^2} I_j$$

where $I_j = 1$ if the j^{th} $t \times t$ submatrix equals J_t [$I_j = 0$ otherwise]. Equation (1) may then be rephrased as

$$\zeta \le n^2 - (t-1)^{1/t} n^{2-\frac{1}{t}} - \frac{n(t-1)}{2} \Rightarrow \mathbf{P}_{u,z}(X=0) = 0 \tag{3}$$

and

$$\zeta \ge \frac{2n^2 \log n}{t}\{1+o(1)\} \Rightarrow \mathbf{P}_{u,z}(X=0) > 0. \tag{4}$$

The rate of growth of the numbers $\zeta(n,t)$ is given by (3) and (4); if $t = n^\alpha$, for example, this rate is of order $n^{2-\alpha}\log n$. We will primarily be concerned with proving results that maintain the flavor of Bollobás' and Znám's results, through the establishment of a threshold phenomenon for $\mathbf{P}_{u,z}(X=0)$, i.e., a threshold function for the bipartite Turán property.

One may obtain a clue as to the direction in which results such as (3) and (4) may be steered by using the following rather elementary probabilistic argument: Suppose that $\mathbf{P}$ denotes the probability measure that independently allots, to each position in $[n]\times[n]$, a one with probability p and a zero with probability $q = 1-p$, where p and q are to be determined. Then, with X representing the same r.v. as before, $\mathbf{E}(X) = \binom{n}{t}^2 p^{t^2} \le K(ne/t)^{2t} p^{t^2}/t \to 0$ if $p = (t/ne)^{2/t}\exp\{(\log t - b_n)/t^2\}$, where $b_n \to \infty$ is arbitrary, so that by Markov's inequality, $\mathbf{P}(X=0) \to 1$ if the *expected* number of ones is less than $n^2(t/ne)^{2/t}\exp\{(\log t - b_n)/t^2\}$. The question, of course, is whether this is true if the *actual* number of ones is at the same level, i.e., under the measure $\mathbf{P}_{u,z}$.

In this paper, we use the extended Janson exponential inequalities [1] to show that both $\mathbf{P}(X=0)$ and $\mathbf{P}_{u,z}(X=0)$ enjoy a sharp threshold at the level suggested by the above reasoning. Specifically, we prove

Theorem. *Consider the probability measure* $\mathbf{P}$ *that independently allots, to each position in* $\mathcal{X} = [n]\times[n] = \{1,2,\ldots,n\}\times\{1,2,\ldots,n\}$, *a one with probability* p *and a zero with probability* $q = 1-p$. *Let* t *satisfy* $\log n \ll t = o(n^{1/3})$, *and set* $X = \sum_{j=1}^{\binom{n}{t}^2} I_j$, *with* $I_j = 1$ *iff* $\mathcal{J} = J_t$, *where* $\mathcal{J}$ *represents the* j^{th} $t\times t$ *submatrix of* $\mathcal{X}$, *and* $I_j = 0$ *otherwise. Then*

$$p = \left(\frac{t}{ne}\right)^{2/t} \exp\left\{\frac{\log t + a_n}{t^2}\right\} \Rightarrow \mathbf{P}(X=0) \to 0 \ \ (n\to\infty)$$

and

$$p = \left(\frac{t}{ne}\right)^{2/t} \exp\left\{\frac{\log t - b_n}{t^2}\right\} \Rightarrow \mathbf{P}(X = 0) \to 1 \ \ (n \to \infty)$$

where $b_n \to \infty$ *is arbitrary, and* $a_n \geq 2t + \log(n^2/t^2) + \delta_n$*, where* $\delta_n \to \infty$ *is arbitrary.*

As a consequence of the above theorem, we will show that it is possible to prove a result with a fixed (as opposed to random) number of ones, i.e., to prove that $\mathbf{P}_{u,z}(X = 0)$ tends to zero or one according as z, the number of ones in the matrix, is larger than $n^2(t/ne)^{2/t}\exp\{(\log t + a_n)/t^2\}$, or smaller than $n^2(t/ne)^{2/t}\exp\{(\log t - b_n)/t^2\}$. This comes as no surprise, since it is well-known that many graph theoretical properties hold under the model $G(n,p)$ if and only if they hold under the model $G(n,m)$, with $m = np$. In particular, with $t = n^\alpha$, we see that J_t submatrices pass from being sparse objects to abundant ones at the level $\zeta = 2(1-\alpha)n^{2-\alpha}\log n$. As a further corollary, we will be able to improve the general upper bound $\zeta(n,t) \leq (2n^2\log n)/t\{1 + o(1)\}$ to $\zeta(n,t) \leq 2n^2(\log(n/t))/t\{1 + o(1)\}$, with the most significant improvement being when $t = n^\alpha$.

The versatility of Janson's inequalities in combinatorial situations has been well-documented; see, for example, the wide range of examples in Chapter 8 of [1] , or the work of Janson, Łuczak, and Ruciński [12], who establish the definitive threshold results for Turán-type properties in the unipartite case. Recent applications of these exponential inequalities include an an analysis of the threshold behaviour of random covering designs ([10]); of random Sidon sequences ([14]); and of the Schur property of random subsets ([13]). A recent analysis of graph-theoretic properties with sharp thresholds may be found in [5].

We end this section by stating the connections between this paper and [9]. In [9], the same problem was treated as in this paper, and the (regular) Janson exponential inequalities yielded the threshold function for the Zarankiewicz property for $2 \leq t \ll n^{1/5}$. A comment was made that the same technique would *probably* work, with a large amount of extra effort, for t's up to $o(n^{1/3})$. In this paper, we choose, instead, to use the *extended* Janson inequalities, together with a different technique for bounding the covariance terms, to prove this fact. We indicate methods by which the main result could, possibly, be extended to $t = o(n^{1/2})$. Other points of difference and similarity with [9] will be indicated at various points throughout this paper.

2. PROOFS

Proof of the Theorem:

We have already provided a proof of the second part of the theorem using nothing more than Markov's inequality, and now turn to the first half. Throughout, we assume that $p = (t/ne)^{2/t}\exp\{(\log t + a_n)/t^2\}$, with conditions on a_n to be determined. Let B_j be the event that the $j^{\rm th}$ $t \times t$ submatrix, denoted by $\mathcal{J}$, equals J_t, i.e., has all ones. We recall the Janson and extended Janson inequalities ([1]):

$$\mathbf{P}(X=0) \le \exp\left\{-\mu + \frac{\Delta}{2(1-\varepsilon)}\right\}; \tag{5}$$

and

$$\mathbf{P}(X=0) \le \exp\left\{-\frac{\mu^2(1-\varepsilon)}{2\Delta}\right\}, \tag{6}$$

where

$$\begin{aligned}\varepsilon =& p^{t^2};\\ \mu =& \binom{n}{t}^2 p^{t^2} = \mathbf{E}(X);\ \text{ and}\\ \Delta =& \mu \sum_{\substack{r,c=1\\ r+c<2t}}^{t} \binom{t}{r}\binom{n-t}{t-r}\binom{t}{c}\binom{n-t}{t-c} p^{t^2-rc};\end{aligned} \tag{7}$$

and (in (6)) provided that $\Delta \ge \mu(1-\varepsilon)$. We also mention the bound based on Chebychev's inequality, known in the combinatorics literature ([1]) as the second-moment bound:

$$\mathbf{P}(X=0) \le \frac{\Delta+\mu}{\mu^2}. \tag{8}$$

In [9], (5) was used to obtain the required threshold for $2 \le t \ll n^{1/5}$ with Δ as in (7), and it was noted that the second moment bound (8) could also be employed–but with a worse rate of approximation, and without any significant reduction in the calculation. It can readily be checked, moreover, that if the *exact* form of (7) is used for Δ, then $\Delta = o(1)$ iff $\mu^2/\Delta \to \infty$ iff $t = o(n^{1/5})$, so that even the extended Janson inequality will not lead to an improvement in the results of [9]. We need, therefore, to work with a different method in conjunction with (6), and proceed as follows: Since $k! \ge A\sqrt{k}(k/e)^k, k = 1, 2, \ldots,$ and

$k! \geq (k/e)^k, k = 0, 1, 2, \ldots$, where $A = e/\sqrt{2}$ and we interpret 0^0 as unity, (7) yields the estimate

$$\Delta \leq \Delta_1 + \Delta_2 \tag{9}$$

where

$$\begin{aligned}\Delta_1 \leq & \frac{4}{e^4}\binom{n}{t}^2 p^{2t^2}\sum_{r,c=1}^{t-1}\left(\frac{te}{c}\right)^c\left(\frac{te}{r}\right)^r\left(\frac{ne}{t-r}\right)^{t-r}\left(\frac{ne}{t-c}\right)^{t-c}\frac{1}{\sqrt{rc(t-r)(t-c)}}p^{-rc}\\ \leq & \binom{n}{t}^2 p^{2t^2}\sum_{r,c=1}^{t-1}\left(\frac{te}{c}\right)^c\left(\frac{te}{r}\right)^r\left(\frac{ne}{t-r}\right)^{t-r}\left(\frac{ne}{t-c}\right)^{t-c}\frac{1}{t-1}p^{-rc}\\ = & \binom{n}{t}^2 p^{2t^2}\sum_{r,c=1}^{t-1}\varphi(r,c) \quad \text{say,}\end{aligned} \tag{10}$$

and

$$\Delta_2 \leq \binom{n}{t}^2 p^{2t^2}\sum_{\substack{\max\{r,c\}=t\\ r+c<2t}}\psi(r,c), \tag{11}$$

where

$$\psi(r,c) = (t-1)\varphi(r,c) = \begin{cases}\left(\frac{te}{c}\right)^c\left(\frac{te}{r}\right)^r\left(\frac{ne}{t-r}\right)^{t-r}\left(\frac{ne}{t-c}\right)^{t-c}p^{-rc} & (\max\{r,c\}<t);\\ e^t\left(\frac{te}{r}\right)^r\left(\frac{ne}{t-r}\right)^{t-r}p^{-rt} & (c=t, r<t);\\ e^t\left(\frac{te}{c}\right)^c\left(\frac{ne}{t-c}\right)^{t-c}p^{-ct} & (r=t, c<t)\\ e^{2t}p^{-t^2} & (r=c=t).\end{cases}$$

Note that φ and ψ are each defined on the compact subset $1,t]^2$ of $\mathbf{R}^2$. Now, in the main result of [9], *both* a_n and b_n could be taken to be arbitrary. We cannot prove such a result, in our current theorem, for t's of the form $\Omega(n^{1/5}) \leq t = o(n^{1/3})$ due, basically, to the above-described "inflation" in the value of Δ. Actually, as we shall see, this is not really an inflation at all: when p equals a slightly higher value, the proof of the theorem will reveal that the maximum summand in Δ (given by (9) through (11)) corresponds to (1,1), whereas the maximum summand in [9]was at $(t-1,t)$, but for a smaller value of p, and with Δ given by (7). The overall effect, however, is for Δ to *decrease*. The proof of the theorem proceeds by a sequence of lemmas:

Lemma 1. *The function $\psi(r,c)$, extended to the closed domain $\mathcal{A} = [1,t]^2 \setminus (t-1,t]^2$ of $\mathbf{R}^2$, has critical points only along the diagonal $\{(r,c) : r = c\}$*

Proof. Writing ψ on the interior of $\mathcal{A}$ as

$$\psi(r,c) = \exp\left\{\log A_c + r\log\left(\frac{te}{r}\right) + (t-r)\log\left(\frac{ne}{t-r}\right) + rc\log s\right\}$$

where A_c depends only on c, and $s = 1/p$, we see that

$$\frac{\partial\psi}{\partial r} = e^{\log\psi}\left\{\log\left(\frac{te}{r}\right) - \log\left(\frac{ne}{t-r}\right) + c\log s\right\}$$

which equals zero if

$$\frac{(t-r)s^c}{r} = \frac{n}{t}.$$

Similarly we verify that $\partial\psi/\partial c = 0$ if $(t-c)s^r/c = n/t$. It follows, that at a critical point,

$$\frac{(t-r)}{rs^r} = \frac{(t-c)}{cs^c}.$$

Now, since the function $\eta(x) = (t-x)/xs^x;\ \ (1 \le x \le t)$, is decreasing, it follows that $\eta(r) = \eta(c) \Rightarrow r = c$. The lemma follows.

Lemma 2. $\psi(1,1) \ge \psi(1,x) = \psi(x,1)\ \ \forall x \in [1,t]$, *provided that* $t^2 = o(n)$ *and* $t \gg \log n$.

Proof. We show that $\psi(1,x)$ is decreasing in x. Since $\psi(1,x) = K(te/x)^x(ne/(t-x))^{t-x}p^{-x}$ for a constant K, we see that the sign of $d\psi(1,x)/dx$ is determined by the quantity $\log(te/x) - \log(ne/(t-x)) + \log s = \log(t(t-x)s/nx)$, which is negative if $t^2 s \le n$. This concludes the proof of Lemma 2, since $p \approx 1$ in all the cases we consider.

Lemma 3. $\psi(1,1) \ge \varphi(1,1) \ge \psi(t,x) = \psi(x,t)\ \forall x \in [1,t-1]$, *provided that* $t^2 = o(n), t \gg \log n$, *and* $p = (t/ne)^{2/t}\exp\{(\log t + a_n)/t^2\}$ *with* a_n *restricted to a range to be specified below.*

Proof. We consider the function $\psi(t,x) = e^t(te/x)^x(ne/(t-x))^{t-x}p^{-tx}$, the sign of whose derivative is determined by the quantity $\log(t(t-x)s^t/nx)$; it is easy to verify that $\psi'(t,x) \ge 0$ provided that $x \le t^2s^t/(n+ts^t)$. We next find conditions under which $t^2s^t/(n+ts^t) \ge t-1$; this inequality may be checked to hold provided that $s^t \ge n$, i.e., if $np^t \le 1$. Now if we set $p = (t/ne)^{2/t}\exp\{(\log t + a_n)/t^2\}$ we see that we must have

$$\exp\{(\log t + a_n)/t\} \le ne^2/t^2 \tag{12}$$

in order for $t^2s^t/(n+ts^t)$ to exceed $t-1$. Since $t^2=o(n)$, we can always choose $a_n\to\infty$ slowly enough so that (12) holds. But we must be more careful, for reasons that will soon become apparent, and note, more specifically, that

$$a_n \le t\log\left(\frac{ne^2}{t^2}\right)-\log t \tag{13}$$

will certainly suffice. Lemma 3 will follow if we can show that $\varphi(1,1)\ge\psi(t,t-1)$, i.e., that $(n/t)^{2t-3}\ge 4p^{-t^2+t}$, and thus, with $p=(t/ne)^{2/t}\exp\{(\log t+a_n)/t^2\}$, that $\exp\{a_n-2t\}\ge Kn/t^2$. The last condition clearly holds if

$$a_n \ge 2t+\log\left(\frac{n}{t^2}\right)+\delta_n, \tag{14}$$

where $\delta_n\to\infty$ is arbitrarily small; since $2t+\log(n/t^2)+\delta_n\le t\log(ne^2/t^2)-\log t$, (13) and (14) complete the proof of Lemma 3.

Lemma 4.

$\varphi(1,1)\ge\max\{\psi(t-1,x): t-1\le x\le t\}$ *under the same conditions as in Lemma 3.*

Proof. Similar to that of Lemma 3; it turns out that Lemma 4 holds if

$$a_n \ge 2t+\log\left(\frac{n^2}{t^2}\right)+\delta_n, \tag{15}$$

for any $\delta_n\to\infty$.

Lemma 5. $\psi(1,1)\ge\psi(r,r)$, *where* (r,r) *is any critical point of* ψ, *provided that* $t=o(n^{1/2})$, *and* $p=(t/ne)^{2/t}\exp\{(\log t+a_n)/t^2\}$, *where* $a_n\le t\log(ne^2/t^2)-\log t$ *is arbitrary.*

Proof. We shall show that $\alpha(r)=\log\sqrt{\psi(r,r)}$, and hence $\beta(r)=\psi(r,r)$, is first decreasing and then increasing as a function of r if a_n is as stated above. Lemma 5 will then follow from Lemma 4. We have $\alpha(r)=r\log(te/r)+(t-r)\log(ne/(t-r))-(r^2/2)\log p$, so that $\alpha(\cdot)$ is increasing whenever

$$\frac{t(t-r)}{nr}\ge p^r. \tag{16}$$

Note that both sides of (16) represent decreasing functions of r, and, moreover, that the left side is convex. We next exhibit the fact that (16) does not hold when $r=1$, but does when $r=t-1$; it will then follow that (16) holds for each $r\ge r_0$.

With $r=1$, (16) is satisfied only if $t^2/n \ge p$, which is clearly untrue since $t^2 = o(n)$ and $p \sim 1$. Let $r = t-1$. (16) is then equivalent to the condition $np^t \le 1$, which may be checked to hold, as in the proof of Lemma 3, for any $a_n \le t\log(ne^2/t^2) - \log t$. This concludes the proof of Lemma 5.

We have proved thus far that the function ψ, and thus the function φ, $[(r,c) \in \{1,2,\ldots,t\}^2 \setminus (t,t)]$, both achieve a maximum at (1,1) provided that t does not grow too rapidly (or too slowly), and that p is large enough, but not too large. Continuing with the proof, we assume that $p = (t/ne)^{2/t}\exp\{(\log t + a_n)/t^2\}$, with $a_n = 2t + \log(n^2/t^2) + \delta_n$, i.e., equal to the value specified by (15). If we can establish that $\mathbf{P}(X=0) \to 0$ with this value of p, then the same conclusion is certainly valid, by monotonicity, if p assumes any larger value. So far, our analysis has led (roughly) to the conditions $\log n \ll t \ll n^{1/2}$; we now see how the "legal" use of Janson's inequalities forces further restrictions on t – in particular, we will need to assume that $\log n \ll t \ll n^{1/3}$. Returning to the extended Janson inequality, we must first find conditions under which $\Delta \ge \mu$; this condition will ensure the validity of (6). Since, by (7), $\Delta \ge K\binom{n}{t}^2 p^{2t^2} t^2 (ne/t)^{2t-2}(1/t)$ for some constant K, and $\mu = \binom{n}{t}^2 p^{t^2}$, we must have

$$Kp^{t^2} \ge \frac{t^{2t-3}}{n^{2t-2}e^{2t-2}}$$

for Δ to exceed μ. Setting $p = (t/ne)^{2/t}\exp\{(a_n + \log t)/t^2\}$, we see that $\Delta \ge \mu$ if

$$K\left(\frac{t}{ne}\right)^{2t} te^{a_n} \ge \frac{t^{2t-3}}{n^{2t-2}e^{2t-2}},$$

i.e., if

$$Kt^4 e^{a_n} \ge n^2 e^2,$$

or, if

$$a_n \ge \log\left(\frac{n^2e^2}{t^4K}\right).$$

This may certainly be assumed to be true, and we next investigate whether we have $\mu^2/\Delta \to \infty$ for $p = (t/ne)^{2/t}\exp\{(a_n + \log t)/t^2\}$; this will be the final step in the proof of the theorem. We have, by Lemmas 1 through 5,

$$\frac{\mu^2}{\Delta} \ge \frac{\binom{n}{t}^4 p^{2t^2}}{t^2\binom{n}{t}^2 p^{2t^2}\varphi(1,1)}$$

$$\begin{aligned}
&= \frac{\binom{n}{t}^2 p}{t^2(te)^2(\frac{ne}{t-1})^{2t-2}(t-1)^{-1}} \\
&\succeq \frac{(n-t)^{2t}p}{t(t/e)^{2t}t^2(te)^2(\frac{ne}{t-1})^{2t-2}(t-1)^{-1}} \\
&\succeq \frac{n^2}{t^6} \to \infty
\end{aligned}$$

if $t = o(n^{1/3})$; in the last two lines of the above calculation, the notation $f \succeq g$ means that $f \geq Kg$ for some positive constant K. This proves the theorem; as in [9], the use of the second moment method would have led to a proof with the same degree of computation as above, but with a far worse approximation for $\mathbf{P}(X = 0)$.

Remarks. Observe that the above proof actually shows, as in [1], pp. 40–41, that $X \sim \mathbf{E}(X)$ with high probability. The condition $t \gg \log n$ arises at several points in our proof and is crucial. In a similar vein, we point out that the condition $t = o(n^{1/3})$ arose at the very end of our proof, when the generalized Janson inequality was invoked. A more careful analysis, using the chain of inequalities $\Delta \leq \binom{n}{t}^2 p^{2t^2}[\varphi(1,1) + t^2T_2]$; $\Delta \leq \binom{n}{t}^2 p^{2t^2}[\varphi(1,1) + T_2 + t^2T_3]$; etc., where $T_2, T_3 \ldots$ represent the second, third,... largest summands in (10) and (11), would clearly lead to improvements. We conjecture, therefore, that the main result is true when $t = o(n^{1/2})$, and also that a_n can be chosen (like b_n) to tend to infinity at an arbitrarily slow rate. The latter fact is known to be true for $t = o(n^{1/5})$ (see [9] for a proof). Now if one seeks to maximize φ (with Δ as in (7)) for $p = (t/ne)^{2/t}\exp\{(a_n + \log t)/t^2\}$, where $a_n \to \infty$ at an arbitrarily *slow* rate, then the maximum is achieved, for all $t = o(n^{1/2})$, at $(t-1, t)$ (see [9]). The problem, however, is that the Janson and extended Janson inequalities are both valid only for $t = o(n^{1/5})$ (as proved in [9]), whilst for a Δ inflated as in (10) and (11), the bound (5) is not useful, and, as we have seen, the extended Janson inequality unfortunately requires, for $t = o(n^{1/3})$, that a_n grow at a *fast enough* rate–with the maximum of φ occurring at (1,1). Graphs of $\varphi(r, c)$, drawn using $\mathcal{MATHEMATICA}^{©}$, show how very sensitive the location of the maximum value of φ is to small changes in the arguments. A new approach is, therefore, needed to resolve the above conjecture. We end with two corollaries:

Corollary 1. *Consider the probability measure $\mathbf{P}_{u,z}$ which uniformly places ζ zeros and $z = n^2 - \zeta$ ones among the entries of the $n \times n$ matrix. Let t satisfy $\log n \ll t = o(n^{1/3})$*

and set $X = \sum_{j=1}^{\binom{n}{t}^2} I_j$*, with* $I_j = 1$ *or* $I_j = 0$ *according as the* j^{th} $t \times t$ *submatrix consists of all ones (or not). Then for any* $b_n \to \infty$*, and* a_n *as in the theorem,*

$$z = n^2 \left(\frac{t}{ne}\right)^{2/t} \exp\left\{\frac{\log t + a_n}{t^2}\right\} \Rightarrow \mathbf{P}_{u,z}(X = 0) \to 0 \ \ (n \to \infty)$$

and

$$z = n^2 \left(\frac{t}{ne}\right)^{2/t} \exp\left\{\frac{\log t - b_n}{t^2}\right\} \Rightarrow \mathbf{P}_{u,z}(X = 0) \to 1 \ \ (n \to \infty)$$

Proof. We clearly have, for each z, $\mathbf{P}_{u,z}(X = 0) = \mathbf{P}(X = 0|\text{the } n \times n \text{ matrix has } z \text{ ones})$. Set $p = (t/ne)^{2/t} \exp\{(\log t + a_n)/t^2\}$ and let z denote the corresponding number of ones. Then

$$\mathbf{P}(X = 0|z = n^2p) \le \mathbf{P}(X = 0|z \le n^2p) \le 3\mathbf{P}(X = 0) \to 0$$

by the theorem, where the last inequality above follows due to the observation that $\mathbf{P}(A|B) \le \mathbf{P}(A)/\mathbf{P}(B)$ and the fact that the central limit theorem [or the approximate and asymptotic equality of the mean and median of a binomial distribution] imply that $\mathbf{P}(z \le n^2p) \ge 1/3$. This proves the first half of the corollary. Conversely, with $p = (t/ne)^{2/t} \exp\{(\log t - b_n)/t^2\}$ the same reasoning implies that

$$\mathbf{P}(X \ge 1|z = n^2p) \le \mathbf{P}(X \ge 1|z \ge n^2p) \le 3\mathbf{P}(X \ge 1) \to 0,$$

again by the theorem. This completes the proof.

Corollary 2. $\zeta(n,t) \le (2n^2/t)(\log(n/t))\{1 + o(1)\}$.

Proof. By Corollary 1,

$$\begin{aligned}
\zeta(n,t) \le & n^2 \left\{1 - \left(\frac{t}{ne}\right)^{2/t} \exp\left\{\frac{\log t - b_n}{t^2}\right\}\right\} \\
= & n^2 \left\{1 - \exp\left\{-\frac{2}{t}\log\left(\frac{ne}{t}\right) + \frac{\log t - b_n}{t^2}\right\}\right\} \\
\le & n^2 \left\{\frac{2}{t}\log\left(\frac{n}{t}\right) + \frac{2}{t} - \frac{\log t}{t^2} + \frac{b_n}{t^2}\right\} \\
= & \frac{2n^2}{t}\log\frac{n}{t}\{1 + o(1)\},
\end{aligned}$$

as asserted.

Acknowledgement

The research of all three authors was partially supported by NSF Grant DMS-9322460. They would like to thank Jerry Griggs and Jianxin Ouyang for introducing them to the Problem of Zarankiewicz.

References

[1] N. ALON AND J. SPENCER, "The Probabilistic Method," John Wiley and Sons, Inc., New York, 1992.

[2] A. BARBOUR, O. CHRYSSAPHINOU, AND M. ROOS, Compound Poisson approximation in systems reliability, *Naval Research Logistics* **43** (1996), 251–264.

[3] A. BARBOUR, L. HOLST, AND S. JANSON, "Poisson Approximation," Clarendon Press, Oxford, 1992.

[4] B. BOLLOBÁS, "Extremal Graph Theory," Academic Press, London, 1978.

[5] E. FRIEDGUT AND G. KALAI, Every monotone graph property has a sharp threshold, *Proc. Amer. Math. Soc.*, to appear (1996).

[6] Z. FÜREDI, New asymptotics for bipartite Turán numbers, *J. Combinatorial Theory, Series A* **75** (1996), 141–144.

[7] Z. FÜREDI, An upper bound on Zarankiewicz' problem, *Comb. Prob. Computing* **5** (1996), 29–33.

[8] C. GENTRY, On the half-half case of the problem of Zarankiewicz. In preparation (1996).

[9] A. GODBOLE AND H. GRAZIANO, Contributions to the problem of Zarankiewicz. Submitted to *J. Graph Theory* (1996).

[10] A. GODBOLE AND S. JANSON, Random covering designs, *J. Combinatorial Theory, Series A* **75** (1996), 85–98.

[11] J. GRIGGS AND J. OUYANG, (0,1)-matrices with no half-half submatrix of ones. To appear in *European J. Comb.*

[12] S. JANSON, T. ŁUCZAK, AND A. RUCIŃSKI, An exponential bound for the probability of non-existence of a specified subgraph in a random graph, *in* " Proceedings

of the 1987 Poznan Conference on Random Graphs," John Wiley and Sons, Inc., New York, 1990.

[13] V. RÖDL AND A. RUCIŃSKI, Rado partition theorem for random subsets of integers. Preprint (1996).

[14] J. SPENCER AND P. TETALI, Sidon sequences with small gaps, *in* "Discrete Probability and Algorithms, IMA Volumes in Mathematics and its Applications," Vol. 72, Springer Verlag, New York, 1995.

[15] Š. ZNÁM, Two improvements of a result concerning a problem of K. Zarankiewicz, *Colloq. Math* **13** (1965), 255-258.