

A NOTE ON THE INDEPENDENCE NUMBER OF TRIANGLE-FREE GRAPHS

James B. SHEARER

Department of Mathematics, University of California, Berkeley, CA 94720, USA

Received 26 February 1982

Revised 16 August 1982

Let G be a triangle-free graph on n points with average degree d . Let α be the independence number of G . In this note we give a simple proof that $\alpha \geq n(d \ln d - d + 1)/(d - 1)^2$. We also consider what happens when G contains a limited number of triangles.

Let G be a triangle-free graph on n vertices with average degree d . Let α be the independence number of G . In [1] Ajtai et al. prove that $\alpha > n \ln d/(100d)$ for $d \geq d_0$. Here we give a simpler proof of a slightly stronger result.

Theorem 1. *Let G be a triangle-free graph on n vertices with average degree d . Let α be the independence number of G . Let $f(d) = (d \ln d - d + 1)/(d - 1)^2$, $f(0) = 1$, $f(1) = \frac{1}{2}$. Then $\alpha \geq nf(d)$.*

Proof. Note f is continuous for $0 \leq d < \infty$ and for $0 < d < \infty$, $1 > f(d) > 0$, $f'(d) < 0$, $f''(d) \geq 0$. Furthermore f satisfies the differential equation

$$(d + 1)f(d) = 1 + (d - d^2)f'(d). \quad (1)$$

We will prove the theorem by induction on n . Note it is true for $n \leq d/f(d)$ as G triangle-free implies the neighbors of any point in G form an independent set so $\alpha \geq d \geq nf(d)$. Let P be a point in G . Let d_1 be the degree of P . Let d_2 be the average degree of the neighbors of P . We claim we may choose P so that

$$(d_1 + 1)f(d) \leq 1 + (dd_1 + d - 2d_1d_2)f'(d). \quad (2)$$

For as P ranges over the points of G the average value of d_1d_2 is equal to the average value of d_1^2 which is $\geq d^2$. Hence the average value of the left hand side of (2) equals the left hand side of (1) while the average value of the right hand side of (2) is \geq the right hand side of (1). Hence (2) holds on the average which suffices to prove the claim.

Let G' be the graph formed from G by deleting P and all its neighbors. Note G' is also triangle-free and contains $n - d_1 - 1$ points and $\frac{1}{2}nd - d_1d_2$ edges. Let d' be the average degree of points in G' . Then $d' = (nd - 2d_1d_2)/(n - d_1 - 1)$. By the induction hypothesis G' contains an independent set of size $(n - d_1 - 1)f(d')$.

Hence by adding P we obtain an independent set in G of size $1 + (n - d_1 - 1)f(d')$. Now since $f''(d) \geq 0$ for $0 < d < \infty$ we have $f(d') \geq f(d) + (d' - d)f'(d)$. Therefore

$$\begin{aligned} 1 + (n - d_1 - 1)f(d') &\geq 1 + (n - d_1 - 1)f(d) + (n - d_1 - 1)(d' - d)f'(d) \\ &= 1 + (n - d_1 - 1)f(d) + (dd_1 + d - 2d_1d_2)f'(d) \\ &\geq (n - d_1 - 1)f(d) + (d_1 + 1)f(d) = nf(d). \end{aligned}$$

Hence we have found an independent set in G of the desired size which completes the proof by induction.

Remark 1. Consider the following algorithm for generating an independent set of points of G . Pick a point P at random and place it in the independent set. Delete P and its neighbors from G and iterate. Since (2) holds on the average, the average size of an independent set generated by this algorithm will be $\geq nf(d)$.

Remark 2. By considering random graphs on n points with average degree d for $n \gg d \gg 1$ we can show the existence of triangle-free graphs on n points with average degree d and independence number

$$\leq n \left[\frac{2 \ln d}{d} - \frac{2 \ln \ln d}{d} + O\left(\frac{1}{d}\right) \right].$$

We now consider what happens when we do not require G to be triangle-free. Given a graph G let $n(G)$ be the number of vertices G contains, $I(G)$ be the independence number of G , $d(G)$ be the average degree of a vertex in G and $T(G)$ be the average number of triangles a vertex of G is contained in. Then define $F(d, T)$ as follows:

$$F(d, T) = \lim_{\substack{n(G) \rightarrow \infty \\ d(G) \rightarrow d \\ T(G) \rightarrow T}} \inf I(G)/n(G) \quad \text{for any } d, T \geq 0. \quad (3)$$

Note Theorem 1 and Remark 2 are equivalent to inequalities

$$f(d) \leq F(d, 0) \leq 2 \ln d/d + O(\ln \ln d/d).$$

Also if G consists of the disjoint union of K_1 copies of G_1 and K_2 copies of G_2 , then

$$n(G) = K_1 n(G_1) + K_2 n(G_2), \quad I(G) = K_1 I(G_1) + K_2 I(G_2)$$

or

$$\begin{aligned} I(G)/n(G) &= [K_1 n(G_1)/n(G)][I(G_1)/n(G_1)] \\ &\quad + [K_2 n(G_2)/n(G)][I(G_2)/n(G_2)], \end{aligned}$$

$$d(G) = \left(\frac{K_1 n(G_1)}{n(G)} \right) d(G_1) + \left(\frac{K_2 n(G_2)}{n(G)} \right) d(G_2) \quad \text{and}$$

$$T(G) = \left(\frac{K_1 n(G_1)}{n(G)} \right) T(G_1) + \left(\frac{K_2 n(G_2)}{n(G)} \right) T(G_2).$$

Hence letting $\rho = K_1 n(G_1)/n(G)$ it follows that

$$\begin{aligned} F(\rho d_1 + (1-\rho)d_2, \rho T_1 + (1-\rho)T_2) \\ \leq \rho F(d_1, T_1) + (1-\rho)F(d_2, T_2), \quad 0 \leq \rho \leq 1. \end{aligned} \quad (4)$$

Using (4) it is easy to show that F is continuous and as a function of d or T alone nonincreasing.

As in [1] we can bound $F(d, T)$ from below in terms of $F(d, 0)$. Let G be a graph on n points with $T(G) \leq 3$. Then G contains $\frac{1}{3}nT(G) \leq n$ triangles. Hence by deleting 1 point from each triangle we obtain a triangle-free induced subgraph G' of G containing at least $(1 - \frac{1}{3}T(G))n$ points. Furthermore $d(G') \leq d(G)/(1 - \frac{1}{3}T(G))$. It follows that

$$F(d, T) \geq (1 - \frac{1}{3}T)F(d/(1 - \frac{1}{3}T), 0). \quad (5)$$

Let G be a graph on n points and let G' be a random induced subgraph on K points. Let $\overline{d(G')}$, $\overline{T(G')}$ be the expected values of $d(G')$ and $T(G')$ respectively. Let $\rho = K/n$. Then

$$\overline{d(G')} = d(G) \frac{K-1}{n-1} \leq \rho d(G) \quad \text{and} \quad \overline{T(G')} = T(G) \frac{(K-1)(K-2)}{(n-1)(n-2)} \leq \rho^2 T(G).$$

It follows from (4) that

$$F(d, T) \geq \rho F(\rho d, \rho^2 T), \quad 0 \leq \rho \leq 1. \quad (6)$$

Combining (5) and (6) we obtain

$$F(d, T) \geq \max_{0 \leq \rho \leq 1} \rho (1 - \frac{1}{3}\rho^2 T) F\left(\frac{\rho d}{1 - \frac{1}{3}\rho^2 T}, 0\right). \quad (7)$$

Let $\varepsilon = \rho^2 T$. Then

$$\begin{aligned} F(d, T) &\geq \rho (1 - \frac{1}{3}\varepsilon) F\left(\frac{\rho d}{1 - \frac{1}{3}\varepsilon}, 0\right) \geq \rho (1 - \frac{1}{3}\varepsilon) f(\rho d/(1 - \frac{1}{3}\varepsilon)) \\ &\geq \rho (1 - \frac{1}{3}\varepsilon) \frac{\ln[\rho d/(1 - \frac{1}{3}\varepsilon)] - 1}{\rho d/(1 - \frac{1}{3}\varepsilon)} \quad \left(\text{using } f(x) \geq \frac{\ln x - 1}{x}\right) \\ &\geq (1 - \frac{2}{3}\varepsilon) \frac{\ln \rho + \ln d - 1}{d} = (1 - \frac{2}{3}\varepsilon) \frac{\ln d - \frac{1}{2}\ln T + \frac{1}{2}\ln \varepsilon - 1}{d} \\ &\geq \frac{\ln d - \frac{1}{2}\ln T - 1}{d} - \frac{2}{3}\varepsilon \frac{\ln d - \frac{1}{2}\ln T - 1}{d} + \frac{\ln \varepsilon}{2d} \end{aligned}$$

(assuming $\varepsilon \leq 1$). This is maximized when

$$\varepsilon = \left[\frac{4}{3}(\ln d - \frac{1}{2}\ln T - 1)\right]^{-1}.$$

However, we must have $\rho \leq 1$ or $\varepsilon \leq T$ or

$$\frac{4}{3}T(\ln d - \frac{1}{2}\ln T - 1) \geq 1 \quad \text{or} \quad T \geq 7/4(\ln d - \frac{1}{2}\ln T - 1).$$

Also $\varepsilon \leq 1 \Rightarrow T \leq e^{-7/2} d^2$. Hence

$$F(d, T) \geq (1 - \frac{2}{3}T) \frac{\ln d - 1}{d} \quad \text{for } 0 \leq T \leq \frac{7}{4(\ln d - \frac{1}{2} \ln T - 1)},$$

$$F(d, T) \geq \frac{\ln d - \frac{1}{2} \ln T - 1}{d} - \frac{\frac{1}{2} \ln \frac{4}{3} e((\ln d - \frac{1}{2} \ln T - 1))}{d}$$

$$\text{for } \frac{7}{4(\ln d - \frac{1}{2} \ln T - 1)} \leq T \leq e^{-7/2} d^2. \quad (8)$$

Note we always have $F(d, T) \geq 1/(d+1)$.

We also can bound $F(d, T)$ from above in terms of $F(d, 0)$ by means of the following construction. Let G be a triangle-free graph on n points and let r be a positive integer. Let G' be the graph obtained from G by replacing each point in G with the complete graph K_r and each edge in G with the complete bipartite graph $K_{r,r}$. Then we have

$$n(G') = rn(G), \quad I(G') = I(G), \quad d(G') = rd(G) + r - 1,$$

$$T(G') = \frac{3}{2}(r)(r-1)d(G) + \frac{1}{2}(r-1)(r-2).$$

Hence we have

$$F(dr + r - 1, \frac{3}{2}dr(r-1) + \frac{1}{2}(r-1)(r-2)) \leq \frac{1}{r} F(d, 0). \quad (9)$$

Since F as a function of d or T alone is non-increasing (9) implies

$$F((d+1)r, \frac{3}{2}(d+1)r^2) \leq \frac{1}{r} F(d, 0). \quad (10)$$

In fact it can be shown (10) holds for non-integral r as well (if $r = S + \theta$, $0 < \theta < 1$ replace $(1 - \theta)n$ points of G with K_S , θn points of G with K_{S+1}). Changing variables in (10) yields

$$F(d, T) \leq \frac{3d}{2T} F\left(\frac{3d^2}{2T} - 1, 0\right), \quad T \leq 3/2 d^2. \quad (11)$$

Combining (11) with Remark 2 we obtain

$$F(d, T) \leq \frac{2 \ln d}{d} + O\left(\frac{1}{d}\right), \quad 0 \leq T \leq d,$$

$$\leq 4 \frac{\ln d - \frac{1}{2} \ln T}{d} + O\left(\frac{1}{d}\right), \quad d \leq T \leq d^2. \quad (12)$$

Remark 3. It follows from (12) that

$$F(d, Ad^2/(\ln d)^2) \leq 4 \frac{\ln d - \ln d + \frac{1}{2} \ln A + \ln \ln d}{d} + O\left(\frac{1}{d}\right)$$

$$= -\frac{4 \ln \ln d}{d} + O(1/d).$$

This shows that Remark 3 in [1] which states in effect that there exist constants A, B such that $F(d, Ad^2/(\ln d)^2) \geq B(\ln d)/d$ for $d \geq d_0$ is incorrect.

Remark 4. Some questions remain. For example what is

$$\lim_{d \rightarrow \infty} dF(d, 0)/\ln d \quad \text{or} \quad \lim_{d \rightarrow \infty} F(d, d)/F(d, 0)?$$

Also what if anything can be proven about the independence number of K_4 -free graphs?

After writing this paper I discovered two other papers [2], [3] which overlap this paper to some extent. In [3] it is shown that the independence number α of K_p -free graphs exceeds $c(n/d) \ln(\ln d/p)$ for $p \geq 4$ but the authors state they are unable to decide whether $\alpha > c_p(n/d) \ln d$ even with $p = 4$.

References

- [1] M. Ajtai, J. Komlós and E. Szemerédi, A dense infinite Sidon sequence, *Europ. J. Combinatorics* 2 (1981) 1–15
- [2] M. Ajtai, J. Komlós and E. Szemerédi, A note on Ramsey numbers, *J. Combin. Theory (A)* 29 (1980) 354–360
- [3] M. Ajtai, P. Erdős, J. Komlós and E. Szemerédi, On Turán's theorem for sparse graphs, *Combinatorica* 1 (1981) 313–317