# A Note on Triangle-Free Graphs

Paul Erdős, Svante Janson, Tomasz Luczak, Joel Spencer

## 1 Introduction

There is a natural, if imprecise, notion that the requirement of trianglefreeness on a graph $G$ forces it into a bipartite-like form. In an extreme case, if $G$ has $n$ vertices and $e = \lfloor \frac{n^2}{4} \rfloor$ edges then Turán's Theorem gives that it must be the bipartite graph $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$. Our concern here is what happens when $e$ is smaller, must $G$ still exhibit bipartite-like behavior. Very roughly, our answer is: Yes, if $e \gg n^{3/2}$, No otherwise.

We define $B(G)$ to be the maximal number of edges over all *induced* bipartite subgraphs of $G$. We define $f(n,e)$ to be $\min B(G)$ over all $G$ with $n$ vertices and $e$ edges. We define $g(e)$ to be $\min B(G)$ over all $G$ with $e$ edges so that $g(e) = \min_n f(n,e)$.

**Theorem 1.**

$$c_1 e^{1/3} \le g(e) \le c_2 e^{1/3} \ln^2 e$$

**Theorem 2.** For $e < c_5 n^{3/2}$

$$c_1 e^{1/3} \le f(n,e) \le c_2 e^{1/3} \ln^2 e$$

while for $c_5 n^{3/2} < e < n^2/4$

$$c_3 e^3 n^{-4} \le f(n,e) \le c_4 e^3 n^{-4} \ln^2 n$$

Here $c_1, c_2 \ldots$ represent absolute positive constant whose optimal values we do not attempt to compute. The polylogarithmic factor between lower and upper bounds would be very nice to eliminate but we have not been able to do so.

## 2 In a paper of Erdős...

The upper bound of Theorem 1 is based on a 1961 paper by the senior author [1]. There it is shown that the random graph $G$ on $n$ vertices and $n^{3/2} A^{-1/2}$ edges ($A$ a sufficiently large constant) almost surely contains a trianglefree graph $G_1$ with no independent set of size $x = An^{1/2} \ln n$. In modern language the random $G(n,p)$ with edge probability $p = A^{-1/2} n^{-1/2}$ almost surely has this property. Let $B^+(G_1)$ be the maximal number of edges of $G_1$ on a subset of $2x$ vertices. Then $B(G_1) \leq B^+(G_1)$ almost surely and $B^+(G_1) \leq B^+(G)$ tautologically. Now

$$\Pr[B^+(G) \geq \alpha] \leq \binom{n}{2x}\binom{\binom{2x}{2}}{\alpha} p^\alpha \leq n^{2x}\left(\frac{2ex^2p}{\alpha}\right)^\alpha = o(1)$$

for $\alpha = cn^{1/2}\ln^2 n$. That is, $B(G_1) = O(n^{1/2}\ln^2 n)$ almost surely. Thus there exist $G_1$ with $n$ vertices, $e = O(n^{3/2})$ edges with

$$B(G_1) = O(n^{1/2}\ln^2 n) = O(e^{1/3}\ln^2 e)$$

as claimed.

# 3 Two Proofs of a Lower Bound

Here we give two arguments that $g(e) = \Omega(e^{1/3})$.

For the first proof we show

$$g(e) \geq \min_{1\leq d\leq e} d + f(e - d(d^2+1))$$

where we interpret $f(a)$ as zero if $a$ is negative. The conclusion $g(e) = \Omega(e^{1/3})$ follows from elementary analysis. Let $G$ be a graph with $e$ edges and let $d$ be the maximal degree $\deg(x)$ over $x \in V(G)$. Let $N(x)$ denote the set of neighbors of $x$, an independent set as $G$ is trianglefree. Consider the restriction of $G$ to $x \cup N(x)$, $B(G) \geq \deg(x) = d$. Let $N^2(x)$ denote the vertices at distance two from $x$. Let $H$ be $G$ with vertices $x \cup N(x) \cup N^2(x)$ deleted. At most $1 + d + d(d-1) = d^2+1$ vertices and hence at most $d(d^2+1)$ edges have been deleted so that $H$ has an induced bipartite $A \times B$ with $f(e - d(d^2+1))$ edges. But then $(A \cup \{x\}) \times (B \cup N(x))$ is an induced bipartite graph with $d + f(e - d(d^2+1))$ edges.

The second proof is probabilistic. Take $c = .1$ for definiteness and we wish to show $g(e) \geq ce^{1/3}$. We can assume $d \leq ce^{1/3}$ since, as already stated, $B(G) \geq d$. Pick a random subset $S$ of the vertices by

$$\Pr[x \in S] = p = ce^{-2/3}$$

¿From $S$ delete all pairs of points at distance one or two, yielding a random set $S^-$. Let $T$ be the set of neighbors of points of $S^-$. ¿From $T$ delete all $y, y'$ for which there is a path $xyy'x'$ of length 3 with $x, x' \in S^-$. Then $G$ restricted to $S^- \cup T^-$ is an induced bipartite graph – indeed, the union of stars centered at the $x \in S^-$. Let $X$ be the number of edges in this random restriction. We claim

$$E[X] = \Omega(e^{1/3})$$

Let us concentrate on a single edge $\{x, y\}$ of $G$. With probability $p(1-p) = \Omega(e^{-2/3})$ we select $x$ but not $y$ in $S$. Now if we don't have $x \in S^-, y \in T^-$ we must have some $x' \neq x, y$ at distance one or two from $x$ or $y$. There are at most $2(d^2+1)$ potential such $x'$ so the probability some such $x' \in S$ is at most $2(d^2+1)p$ which is certainly less than .01 by our choice of $c$. With probability at least .99 this does not occur. Thus the probability that $x \in S^-, y \in T^-$ is still $\Omega(e^{-2/3})$. The bound on $E[X]$ follows by the linearity of expectation. Some choice must reach the expectation so there exist $S, T$ and hence $S^-, T^-$ with $X \geq c_2 e^{1/3}$ as desired.

## 4 $n$ vertices, $e$ edges

Here we show Theorem 2. Suppose $e \leq c_5 n^{3/2}$. Clearly $f(n,e) \geq g(e) = \Omega(e^{1/3})$. The construction of Theorem 1 has $e = \Theta(n^{3/2})$. For $e \leq c_5 n^{3/2}$ there is such a constuction with $n' = \Theta(e^{3/2}) \leq n$ vertices. Adding isolated vertices if necessary gives a graph $G$ with $n$ vertices, $e$ edges and $B(G) = O(e^{1/3} \ln^2 e)$.

Now suppose $e > c_5 n^{3/2}$ and we show $f(n,e) = O(e^3 n^{-4} \ln^2 n)$. As $f(n,e) \leq \binom{n}{2}$ tautologically we can further assume $e = o(n^2)$. Set $m = e^2 n^{-3}$ and split the $n$ vertices in disjoint sets $I_i$, $1 \leq i \leq \frac{n}{m}$, each of size $m$. There is a graph $G_0$ on vertex set $1, \ldots, \frac{n}{m}$ with $\Theta((n/m)^{3/2})$ edges and $B(G_1) = O((n/m)^{1/2} \ln^2(n/m))$. Define $G$ on $n$ vertices by blowing up point $i$ into set $I_i$. That is, the vertex set is the union of the $I_i$ and $x \in I_i, y \in I_j$ are joined if and only if $i, j$ are adjacent in $G_0$. Then $G$ is trianglefree and has $\Theta(m^2(n/m)^{3/2}) = \Theta(e)$ edges with

$$B(G) = O(m^2(n/m)^{1/2} \ln^2(n/m)) = O(e^3 n^{-4} \ln^2(n/m)) = O(e^3 n^{-4} \ln^2 n)$$

as desired.

Finally, we turn to the lower bound. Set $d = 2e/n$, the average degree. Remove sequentially vertices of degree at most, say, $d/10$. This can remove at most $nd/10 = e/5$ edges so one is left with an induced graph $G'$ on $n' \leq n$ vertices and at least $.8e$ edges, all degrees at least $d/10$. Now, restricting our attention to $G'$ we find an induced subgraph of the form $N(x) \times N(y)$. Note that since $G'$ is trianglefree this graph will be an induced bipartite graph. (If the sets intersect we consider it as $N(x) \times N(y) - N(x)$.) Set $f(x, y)$ equal to the number of edges in $N(x) \times N(y)$. Observe that $f(x, y)$ is precisely the number of paths of length three from $x$ to $y$ since each path $xaby$ corresponds to a unique edge $ab$. Thus $\sum f(x, y)$ gives the number of ordered 3-paths in $G'$. But this number is at least $(1.6e)(.1d - 1)^2 = \Omega(ed^2)$ since we can choose the ordered edge $ab$ is at least $1.6e$ ways and then neighbors $x$ of $a$ and $y$ of $b$. Therefore $\sum f(x, y) = \Omega(ed^2)$. We deduce that for the average $x, y$ and hence for some $x, y$

$$f(x, y) = \Omega(ed^2/(n')^2) = \Omega(ed^2/n^2) = \Omega(e^3 n^{-4})$$

as desired.

## References


[1] P. Erdős, Graph Theory and Probability II, Canad. J. Math *13 (1961), 346-352*