Studia Scientiarum Mathematicarum Hungarica 27 (1992), 403–407

GRAPHS WITH MAXIMUM NUMBER OF STAR-FORESTS

Z. FÜREDI

Abstract

Let **H** denote the vertex disjoint union of stars of $a_1, \dots, a_t$ edges. Here it is proved that if $a_i > \log_2(t+1)$ for all $1 \leqq i \leqq t$ and e is sufficiently large ($e > e_0(a_1, \dots, a_t)$), then a star-forest of e edges and t components contains the largest number of (not necessarily induced) copies of **H**. A simple construction shows that the constraint $a_i = \Omega(\log t)$ cannot be omitted.

This (partly) settles a conjecture of Noga Alon.

1. Notations, preliminaries

Let **G** and **H** be simple graphs (i.e. undirected, finite, no loops and multiple edges) without isolated vertices. In this paper we investigate $N(\mathbf{G}, \mathbf{H})$, the number of subgraphs of **G** isomorphic to **H**. For simplicity, we suppose that the *edges* of the graphs are labelled, so, e.g., $N(\mathbf{K}^n, \mathbf{K}^m) = n(n-1)\dots(n-m+1)$. Let

$$N(e, \mathbf{H}) = \max\{N(\mathbf{G}, \mathbf{H}) : |E(\mathbf{G})| = e\},$$

the maximum number of ways as **H** can be embedded as a subgraph. **G** is called maximal with respect to **H** if $N(\mathbf{G}, \mathbf{H}) = N(|E(\mathbf{G})|, \mathbf{H})$.

A *star* $\mathbf{H}(a)$ is a graph of a edges, $a+1$ vertices with a degree a. The vertex disjoint union of $\mathbf{H}(a_1), \dots, \mathbf{H}(a_t)$ is denoted by $\mathbf{H}(a_1, \dots, a_t)$, and called a star-forest of type $(a_1, \dots, a_t)$. The vector $(a_1, \dots, a_t)$ is abbreviated as $\mathbf{a}$. In this paper we always suppose that $a_i \geqq 2$ for all i, and that $t \geqq 2$, except if otherwise stated.

Alon [1] determined the order of magnitude of $N(e, \mathbf{H})$ whenever **H** is an arbitrary given graph and $e \to \infty$.

Conjecture 1.1 (Alon [2]). *If* $\mathbf{H}(\mathbf{a})$ *is a star-forest and* **G** *is maximal with respect to* **H**, *then* **G** *is a star-forest, too.*

He proved the case $t \leqq 2$. The aim of this paper is to prove 1.1 for a large class of additional cases.

This research was supported in part by the Hungarian National Science Foundation under Grant No. 1812.

Denote the polynomial

$$\sum x_{i_1}^{a_i} x_{i_2}^{a_2} \cdots x_{i_t}^{a_t}$$

by $p(a_1, \dots, a_t, x_1, \dots, x_n)$ or briefly by $p(\mathbf{a},\mathbf{x})$, where $i_1, \dots, i_t$ run over all the $n(n-1)\cdots(n-t+1)$ ordered t-tuples of $\{1, 2, \dots, n\}$. Let $p(\mathbf{a}, n)$ denote

$$\max\Big\{p(\mathbf{a},\mathbf{x}) : \mathbf{x} \geqq 0 \text{ and } \sum_{i=1}^{n} x_i = 1\Big\}.$$

Finally, let $p(\mathbf{a}) = \sup_{n \geqq t} p(\mathbf{a}, n)$.

During the proof $\varepsilon_1, \varepsilon_2, \dots$ and $c_1, c_2, \dots$ denote (explicitly computable) positive constants depending only on $\mathbf{a}$.

2. An asymptotic result

THEOREM 2.1. *Suppose that $a_i \geqq 2$ for all i and $\sum a_i = A$. Then $N(e, \mathbf{H}(\mathbf{a})) = p(\mathbf{a})e^A + O(e^{A-1})$, as e tends to infinity.*

PROOF. First we show that for some $n_0 = n_0(\mathbf{a})$ one has $p(\mathbf{a}, n_0) = p(\mathbf{a}, n)$ for all $n > n_0$, whenever all $a_i \geqq 2$. Suppose that $\mathbf{x}$ is a maximum point with $\mathbf{x} > 0$. Lagrange's multiplicator method gives that

$$\frac{\partial p(\mathbf{a},\mathbf{x})}{\partial x_j} = \lambda \tag{2.1}$$

for all $1 \leqq j \leqq n$. As every term in the polynomial $(\partial/\partial x_j)p(\mathbf{a},\mathbf{x})$ has degree $A - 1$ and has a factor x_j (by $a_i \geqq 2$) we obtain

$$\frac{\lambda}{x_j} = \frac{1}{x_j}\frac{\partial}{\partial x_j} p(\mathbf{a},\mathbf{x}) \leqq (\max a_i - 1)\Big(\sum x_i\Big)^{A-2} = \max_i a_i - 1$$

implying

$$\lambda \leqq x_j(\max a_i - 1). \tag{2.2}$$

On the other hand, summing λx_j for all j (2.1) gives a lower bound for λ

$$\begin{aligned}\lambda = \lambda\Big(\sum x_j\Big) = \sum x_j \frac{\partial}{\partial x_j} p(\mathbf{a},\mathbf{x}) = (A-t)p(\mathbf{a},\mathbf{x}) \geqq \\ \geqq (A-t)p\Big(\mathbf{a}, \Big(\frac{1}{t}, \dots, \frac{1}{t}, 0, 0, \dots, 0\Big)\Big) = (A-t)\frac{t!}{t^A}.\end{aligned} \tag{2.3}$$

If $n \geqq t^A/t!$ and $x_j \leqq 1/n$, then (2.2) and (2.3) contradict each other.

Suppose that $p(\mathbf{a}) = p(\mathbf{a}, x_1, \dots, x_n)$ where $\mathbf{x} \geqq 0$, $\sum x_i = 1$. Then the graph $\mathbf{H}(\lfloor x_1 e \rfloor, \dots, \lfloor x_n e \rfloor)$ contains $p(\mathbf{a})e^A - O(e^{A-1})$ copies of $\mathbf{H}(\mathbf{a})$.

To prove the upper bound, consider an $\mathbf{H}(\mathbf{a})$-maximal graph $\mathbf{G}$ of e edges. First we claim that there is a set $C \subset V(\mathbf{G})$, $|C| \leqq c_1$ $(= c_1(\mathbf{a}))$, such that C intersects all edges of $\mathbf{G}$ and

$$\deg_G(v) \geqq \varepsilon_1 e \tag{2.4}$$

holds for all $v \in C$ for some $\varepsilon_1 = \varepsilon_1(\mathbf{a}) > 0$.

For an edge $E \in E(\mathbf{G})$ denote its *multiplicity* by $M(E)$, i.e. the number of occasions it appears in a subgraph of $\mathbf{G}$ isomorphic to $\mathbf{H}(\mathbf{a})$. Set $M_{\max} = \max\{M(E) : E \in E(\mathbf{G})\}$, and let $\{u, v\} \in E(\mathbf{G})$ be an edge with maximal multiplicity, $M(\{u, v\}) = M_{\max}$. As $p(\mathbf{a}) \geqq t^{-A}$ we have that

$$M_{\max} > \varepsilon_2 e^{A-1} \tag{2.5}$$

holds (for all $e \geqq A$).

Consider an arbitrary edge $\{p, q\} \in E(\mathbf{G})$ and suppose that $M(\{p, q\}) < \frac{1}{3} M_{\max}$. At least $\frac{2}{3} M_{\max}$ copies of $\mathbf{H}(\mathbf{a})$ contains $\{u, v\}$ but not $\{p, q\}$. At least half of these (i.e. $\geqq M/3$) has u as a center of a star. Then delete $\{p, q\}$ from $\mathbf{G}$ and add a new edge $\{u, w\}$ where $w \notin V(\mathbf{G})$. This operation increased $N(\mathbf{G}, \mathbf{H}(\mathbf{a}))$, a contradiction. We obtained that

$$M(E) > \varepsilon_3 e^{A-1} \tag{2.6}$$

holds for all edges $E \in E(\mathbf{G})$. Denote the degrees of the end points of the edge E by d_1, d_2, and let $d = \max\{d_1, d_2\}$. Then E is contained in at most

$$\sum_{\alpha=1,2} \sum_i \binom{d_\alpha}{a_i - 1} a^{A-a_i} < 2ta^{A-1} \left(\frac{d}{e}\right)^{\min a_i - 1} \leqq 2tde^{A-2}$$

star-forests of $\mathbf{G}$. Then (2.6) implies that at least one end point of E must have degree at least $(\varepsilon_3/2t)e$, yielding (2.4).

Finally, let $\mathbf{G}'$ be the bipartite graph obtained by deleting all edges inside $C, C = \{v_1, \dots, v_n\}$, $(n \leqq c_1)$. We get

$$N(\mathbf{G}, \mathbf{H}(\mathbf{a})) \leqq N(\mathbf{G}'(\mathbf{a}), \mathbf{H}) + \binom{c_1}{2}(A - t)e^{A-1}.$$

It is quite clear that for $x_i := \deg_{\mathbf{G}'}(v_i)/e$ one has

$$N(\mathbf{G}', \mathbf{H}(\mathbf{a})) \leqq p(\mathbf{a}, \mathbf{x})e^A + O(e^{A-1})$$

yielding the desired upper bound

$$N(\mathbf{G}, \mathbf{H}(\mathbf{a})) \leqq p(\mathbf{a}, \mathbf{x})e^A + O(e^{A-1}). \tag{2.7}$$

3. An exact statement

THEOREM 3.1. *Suppose that $a_i > \log_2(t+1)$ for all $1 \leqq i \leqq t$ and $\mathbf{G}$ is an $\mathbf{H}(\mathbf{a})$-maximal graph with e edges. If $e > e_0(\mathbf{a})$, then $\mathbf{G}$ is the union of t stars.*

This is not true in general. E.g., if $\mathbf{a} = (a, a, \dots, a)$, then

$$p(\mathbf{a}(x_1, \dots, x_t)) = \frac{t!}{t^{at}} < p(\mathbf{a}, (x_1, \dots, x_{t+1})) = \frac{(t+1)!}{(t+1)^{at}}$$

whenever $a \leqq \ln(t+1)$.

The main tool of the proof of 3.1 is the following technical lemma about $p(\mathbf{a},\mathbf{x})$. This lemma will be proved in the next section.

LEMMA 3.2. *Suppose that $a_i > \log_2(t+1)$ for all $1 \leqq i \leqq t$, $A = \sum a_i$. Suppose further that $x_1, \dots, x_n \geqq \varepsilon$ where $n > t$. Then*

$$p(\mathbf{a}) \geqq p(\mathbf{a},\mathbf{x}) + \frac{\varepsilon^A}{t}.$$

PROOF OF THEOREM 3.1. As we have seen in (2.4), there is a set $C = \{v_1, \dots, v_n\} \subset V(\mathbf{G})$ of large degrees ($\geqq \varepsilon_1 e$). Denote the degree sequence of C by $x_1 e, \dots, x_n e$. Then (2.7) implies that $|p(\mathbf{a},\mathbf{x}) - p(\mathbf{a})| = O(1/e)$. Then Lemma 3.2 gives that $n = t$.

There is no edge outside C, so each component of $\mathbf{H}(\mathbf{a})$ must intersect C. Hence each edge inside C has multiplicity 0, that is, C does not contain any edge by (2.6). Finally, it is clear that all vertices outside C must be of degree exactly one.

4. The proof of Lemma 3.2

Suppose that $x_1 \geqq x_2 \geqq \dots \geqq x_n \geqq \varepsilon$. Denote the sum of all terms of $p(\mathbf{a},\mathbf{x})$ containing x_i by p_i, and let $p_{n-1,n}$ denote the sum of terms containing both x_{n-1}, x_n. As x_n is the smallest of the x_i we have that $p_n \leqq (t/n)p(\mathbf{a},\mathbf{x})$. Similarly, as x_{n-1} is the second smallest of the x_i we obtain that

$$p_{n-1,n} \leqq \frac{t-1}{n-1} p_n \leqq \frac{t-1}{t} p_n. \tag{4.1}$$

Consider the ratio of the sum of distinct terms in p_n and p_{n-1} and use (4.1). We obtain

$$p_{n-1} - p_{n-1,n} \geqq (p_n - p_{n-1,n})\left(\frac{x_{n-1}}{x_n}\right)^a \geqq \frac{p_n}{t}\left(\frac{x_{n-1}}{x_n}\right)^a, \tag{4.2}$$

where $a = \min a_i$. Now define

$$y_i = \begin{cases} x_i & \text{for } i = 1, 2, \dots, n-2 \\ x_{n-1} + x_n & \text{for } i = n-1 \\ 0 & \text{for } i = n. \end{cases}$$

Consider $p(\mathbf{a},\mathbf{y}) - p(\mathbf{a},\mathbf{x})$. We have that the increase of p is at least

$$(4.3) \qquad (-p_n - p_{n-1} + p_{n-1,n}) + (p_{n-1} - p_{n-1,n})\left(\frac{x_n + x_{n-1}}{x_{n-1}}\right)^a.$$

Using (4.2) we have that the expression in (4.3) is at least

$$-p_n + \frac{p_n}{t}\left(\frac{x_{n-1}}{x_n}\right)^a\left(\left(\frac{x_n + x_{n-1}}{x_{n-1}}\right)^a - 1\right).$$

Here the coefficient of p_n/t is $(1+c)^a - c^a$ where $c = x_{n-1}/x_n \geqq 1$. So this coefficient is at least $2^a - 1 \geqq t + 1$. This implies that

$$p(\mathbf{a}) \geqq p(\mathbf{a}, n-1) \geqq p(\mathbf{a},\mathbf{y}) \geqq p(\mathbf{a},\mathbf{x}) + \frac{p_n}{t} \geqq p(\mathbf{a},\mathbf{x}) + \frac{\varepsilon^A}{t}.$$

Remarks, problems

It is probably not too difficult to give an asymptotic formula like in Theorem 2.1 for all $\mathbf{H}(\mathbf{a})$, when some $a_i = 1$ appear.

Another step to prove Conjecture 1.1 would be to get rid of the constraint $a_i > \log_2(t+1)$ in Theorem 3.1. It is easy to prove that if all $a_i \geqq 3$, then in a $\mathbf{H}(\mathbf{a})$-maximal $\mathbf{G}$ all the vertices outside C (see (2.4)) have degree 1.

It also seems to me a solvable question to investigate $N(\mathbf{G},\mathbf{H})$ where now $\mathbf{G}$ and $\mathbf{H}$ are *multigraphs*.

REFERENCES

[1] ALON, N., On the number of subgraphs of prescribed type of graphs with a given number of edges, *Israel J. Math.* **38** (1981), 116–130. *MR* **82b**: 05078

[2] ALON, N., On the number of certain subgraphs contained in graphs with a given number of edges, *Israel J. Math.* **53** (1986), 97–120. *MR* **87j**: 05090

(Received February, 1989, in revised form September, 1990)

MTA MATEMATIKAI KUTATÓINTÉZETE
POSTAFIÓK 127
H–1364 BUDAPEST
HUNGARY