ON THE MAXIMUM SIZE OF RANDOM TREES

M. Protasi*, M. Talamo**

* Dipartimento di Matematica, Università dell'Aquila, Via Roma 33, 67100 L'Aquila. Incaricato di ricerca presso lo IASI-CNR.

** Istituto di Analisi dei Sistemi ed Informatica del C.N.R., Viale Manzoni 30, 00185 Roma.

ABSTRACT

In this paper we prove a conjecture of Erdös and Palka on the maximum size of random trees. Furthermore, while, generally speaking, in the probabilistic analysis the results are proved only when the size of the graphs tends to infinity, in this case, with extremely small probability of error, the results also hold for graphs of small size.

1. INTRODUCTION

The problem of finding the size of the largest induced tree in a graph is NP-complete (see, for instance, [4]) and because of its computational intractability has been studied in the last years using alternative approaches to the classical worst case analysis. In particular, some research effort has been spent analyzing this problem on random graphs, with the aim of evaluating the optimal solution and finding very good approximate solutions for almost all graphs. The achieved results can be summarized as follows. If we assume the constant density model, after some first results in [5], in [1] and in [6], independently, it has been shown that the size of the largest tree is about $\frac{2 \log n}{\text{constant}}$ for almost all graphs of size n. Furthermore in [6], it has been proved that a simple greedy algorithm achieves a solution whose value is one half of the value of the optimal solution. As above written, these results are stated in the constant density model, that is we assume a constant probability p of existence of an edge between two nodes. The situation is different if we are in a more general setting.

In [1], Erdös and Palka posed the following open problem. Let now p be a function on n, i.e. $p = p(n)$ with p tending to zero as n tends to infinity. Find such a value of the edge probability p for which a random graph has the largest induced tree. They conjectured that for a suitable p(n) a random graph contains a tree of size b(c)•n, where b(c) depends only on the constant c. In this paper we prove the conjecture evaluating a lower and upper bound to the size of the tree. In order to achieve this result we will exploit a general probabilistic model introduced in [8] and [9] for studying the max clique and the max independent set problems on random graphs. This general model allows to study graphs of different density and in particular sparse graphs as in the case of the open problem posed by Erdös.

The other important point of the paper regards the asymptoticity of the analysis. In fact with a probability of error tending to zero very fast, we are able to state our results not only, as generally happens, when the number of the nodes of the graph is sufficiently large, but also for graphs of small size. This characteristics, of course, has interesting practical consequences because in the applications we are interested in graphs having a reasonable number of nodes. The possibility of performing such a study has due substantially to the properties of the model we use and in [9] other non asymptotic results have been found.

2. THE PROBLEM

First of all, we recall the problem we want to study. Given a graph, we want to find the largest induced tree, that is, the induced tree having the maximum number of nodes. As we said in the introduction, the problem has been already solved from a probabilistic point of view if we assume a constant probability p ($0 < p < 1$) of existence of an edge between two nodes independently from the presence or absence of any other edges. With this model the number of nodes of the largest induced tree is $2 \log n/[\log(1/(1-p)]+o(\log n)$ for almost all graphs of size n. However, assuming a constant density model, we are able to deal only with dense graphs.

The probabilistic analysis of our problem becomes much more interesting and complete if we also include the case of the sparse graphs. It is natural to expect that in this case we are able to find trees of larger cardinality than for dense graphs. This situation has lead Erdös and Palka [1] to formulate the following open problem:
Let now p be a function on n, i.e. $p = p(n)$ with p(n) tending to zero as n tends to infinity. Find such a value of the edge probability p for which a random graph has the largest induced tree.

Erdös and Palka conjectured that for a suitable p(n) a random graph contains a tree of size b(c)•n, where b(c) depends only on c and c is a constant.

In order to prove, in the following paragraph, that the conjecture is true we will use a general probabilistic model already used in [8] and [9].

DEFINITION 1. Let V be a set of n nodes. Every pair (i,j) with $i,j \in V$ is an edge with probability $p(n) = 1-c^{-c/n}$ independently from the presence or absence of any other edges.

The type of graphs we study depends, of course, on the value of c we choose. Therefore if we allow that c can assume values belonging to a large range, we are able to perform a very general analysis. In particular, for suitable values of c, we are able to deal, at the same time, with dense and sparse graphs. To prove the conjecture we will have to find particular values of c such that the corresponding values of p(n) make the random graph have induced trees as large as possible.

3. MAIN RESULTS

Before giving the main results, we need to state some preliminary lemmas.

Let z_n be the value of the maximum induced tree of a graph of n nodes.

In order to bound some combinatorial inequalities more easily, in the following instead of z_n, we will often use the formula θn.

LEMMA 1.

$$1-c^{-\frac{c}{n}} = \frac{c\ln c}{n} \quad \text{as } n \to \infty$$

PROOF. It is trivial applying some elementary analytic steps.

LEMMA 2. For any fixed $\varepsilon > 0$ and $c \geq e$ constant

$$z_n > \left(\frac{1}{c} - \varepsilon\right)\frac{\log\log n}{\log n} \cdot n$$

for a suitable probability p(n).

PROOF. See [7].

LEMMA 3.

$$\binom{n}{\theta n} \leq \left(\theta^{\theta}(1-\theta)^{(1-\theta)}\right)^{-n}$$

PROOF. See [3] or [9].

Let us now prove the conjecture

THEOREM 1. $0{,}42n \leq z_n \leq 0{,}91n$ almost surely.

PROOF. Let $k = \theta n$. Given a tree T, let $|T|$ denote the number of nodes of T. The proof is divided in two parts. First of all we will find an upper bound to z_n.

1) $\mathrm{Prob}(\exists T / |T| \geq k) \leq \binom{n}{k} \cdot k^{k-2} \cdot p^{k-1} \cdot q^{\binom{k}{2}-(k-1)}$ (where $q = 1-p$)

(Since $k = \theta n$ and $p = 1-c^{-\frac{c}{n}}$)

$$\leq \binom{n}{\theta n} \cdot k^{k-2} (1-c^{-\frac{c}{n}})^{k-1} \cdot (c^{-\frac{c}{n}})^{\frac{(k-1)(k-2)}{2}}$$

(Exploiting Lemmas 1 and 3)

$$\leq (\theta^{\theta} (1-\theta)^{(1-\theta)})^{-n} \cdot (\theta n)^{\theta n-2} (\frac{c\ln c}{n})^{\theta n-1} \cdot (c^{-\frac{c}{n}})^{\frac{(\theta n-1)(\theta n-2)}{2}}$$

$$\leq (\theta(1-\theta)^{\frac{1}{\theta}-1})^{-\theta n} \cdot \theta^{\theta n-2} \cdot n(c\ln c)^{\theta n-1} \cdot c^{-\frac{\theta c}{2}(\theta n-1)}$$

(Exploiting Lemma 2)

$$\leq ((\theta(1-\theta)^{\frac{1}{\theta}-1})^{-1} \cdot \theta c\ln c \cdot c^{-\frac{\theta c}{2}})^{\theta n-1} = \left(\frac{c^{(1-\frac{\theta c}{2})} \ln c}{(1-\theta)^{(\frac{1}{\theta}-1)}} \right)^{\theta n-1}$$

This quantity tends to zero if $\frac{c^{(1-\theta)} \cdot \ln c}{(1-\theta)^{(\frac{1}{\theta}-1)}} < 1$. If we succeed, among the possible values of c that verify the last inequality, in finding the value that maximes θn, we will be immediately able to evaluate an upper bound to z_n. We study the threshold

$$\frac{c^{(1-\frac{\theta c}{2})} \ln c}{(1-\theta)^{(\frac{1}{\theta}-1)}} = 1$$

Applying numerical methods we obtain that the maximum value of θ such that θn is the value of the largest induced tree is less than 0,91.

2) In order to find the lower bound to z_n and complete the proof of the proof of the conjecture, we consider the following inequality. Let y_k be the random variable that denotes the number of trees of size < k

$$\text{Prob}(y_k = 0) \leq \frac{\sigma^2(n,k)}{E^2(n,k)}$$

where σ^2 is the variance and E is the expectation of y_k.
Now we bound the ratio $\frac{\sigma^2(n,k)}{E^2(n,k)}$

$$\frac{\sigma^2(n,k)}{E^2(n,k)} \leq \frac{\sum_{l=1}^{k} k^{k-2}(k-1)^{k-l-2} k^{k-l-1}(k-l+1)^{l}\binom{k}{l}\binom{n}{k}\cdot\binom{n-k}{k-l}\left(\frac{c\ln c}{n}\right)^{2k-4-l} c^{-\frac{c}{n}\left(2\binom{k}{2}-2k-\binom{l}{2}\right)}}{\binom{n}{k}^2 k^{2k-4}\left(\frac{c\ln c}{n}\right)^{2k-4} c^{-\frac{c}{n}2\binom{k}{2}-2k+4}}$$

(By doing some combinatorial and algebraic steps and remembering that $k = \theta n$)

$$\leq \frac{\left(\frac{\theta}{1-\theta}\right)^{\theta n}\left(\frac{n}{c\ln c}\right)^{\theta n} c^{\frac{c\theta^2 n}{2}}}{(\theta n)^{\theta n}} \cdot n^2$$

$$\leq \left(\frac{\left(\frac{\theta}{1-\theta}\right) c^{\frac{c\theta}{2}}}{\theta c\ln c}\right)^{\theta n}(1+\varepsilon) \qquad (\varepsilon > 0)$$

This quantity tends to zero if $\theta \leq 0,42$.

QED

After proving the conjecture, we will give another theorem that shows that the result also hold for graphs pf small size, so strengthening remarkably the achieved result.

Let p_e be the probability of error that z_n does not belong to the interval $[0,42n, 0,91n]$.

THEOREM 2. $p_e < d^n$ for every n, where d is a constant < 1.

Therefore Theorem 2 assures that, with a very small probability of error, that is, with a probability of error tending to zero very fast, the result of Theorem 1 holds for graphs of every size.

4. REFERENCES

[1] P. Erdös, Z. Palka: *Trees in random graphs*, Discr. Math., Vol. 46 (1983).

[2] P. Erdös, A. Renyi: *On the evolution of random graphs*, Publ. Math. Inst. Hung. Acad. Sci., Vol. 5A (1960).

[3] J. Friedman: *Constructing* 0(n log n) *size monotone formulae for the* k-th *elementary symmetric polynomial of* n *boolean variables*, Proc. 25th Symp. on Foundations of Computer Science (1984).

[4] M.R. Garey, D.S. Johnson: *Computers and intractability. A guide to the theory of NP completeness*, Freeman, San Francisco (1978).

[5] M. Karonski, Z. Palka: *On the size of a maximal induced tree in a random graph*, Math. Slovaca, Vol. 30 (1980).

[6] A. Marchetti Spaccamela, M. Protasi: *The largest tree in a random graph*, Theor. Comp. Sci., Vol. 23 (1983).

[7] Z. Palka, A. Rucinski: *On the order of the largest induced tree in a random graph*, (to appear).

[8] M. Protasi, M. Talamo: *A new probabilistic model for the study of algorithmic properties of random graph problems*, Proc. Conference on Foundations of Computation Theory, Borgholm, Lect. Notes in Comp. Sci., Springer Verlag (1983).

[9] M. Protasi, M. Talamo: *A general analysis of the max independent set and related problems on random graphs*, Tech. Rep. n. 3/84, Dipartimento di Matematica, Università del-l'Aquila (1984).