INFORMATIQUE THÉORIQUE ET APPLICATIONS

M. PROTASI
M. TALAMO

A parametric analysis of the largest induced tree problem in random graphs

Informatique théorique et applications, tome 20, nº 3 (1986), p. 211-219
<http://www.numdam.org/item?id=ITA_1986__20_3_211_0>

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 20, n° 3, 1986, p. 211 à 219)

A PARAMETRIC ANALYSIS OF THE LARGEST INDUCED TREE PROBLEM IN RANDOM GRAPHS (*)

by M. Protasi [1] and M. Talamo [2]

Communicated by G. Ausiello

Abstract. – *The optimal value of the largest induced tree is evaluated for a wide family of random graphs. Using this analysis it is possible to solve a conjecture proposed in this area by Erdös and Palka.*

Résumé. – *On présente l'évaluation de l'arbre induit maximal pour une large famille de graphes aléatoire. En utilisant cette analyse on parvient a démontrer une conjecture qui avait été proposée par Erdös et Palka.*

INTRODUCTION

The problem of finding the induced tree with the maximum number of nodes (shortly the largest tree) in a graph has been extensively studied in the last years from a combinatorial and algorithmic point of view. Many efforts have been devoted to this problem in the case of random graphs. In this particular setting some interesting results have been achieved but remarkable questions have to be answered. The aim of this paper is to study the problem in a more general probabilistic framework with respect to the results already given in the literature. From our analysis in particular we will be able to solve a conjecture posed by Erdös and Palka.

In order to present our analysis, first of all we review some preceding results about this problem. If we assume the constant density model (that

(*) Received in May 1984, revised in December 1985.

([1]) Dipartimento di Matematica, Università dell'Aquila, 67100 L'Aquila, Italy.
([2]) Istituto di Analisi dei Sistemi ed Informatica del C.N.R., Viale Manzoni 30, 00185 Roma, Italy.

0296-1598 86/03 211 9/$2.90/© Gauthier-Villars

is we assume random graphs with constant edge probability) in [1] and, independently, in [4], it was show, that the size of the largest induced tree in a random graph of n nodes is about $2 \log n/\text{constant}$. Some partial results can be also found in [3]. Furthermore in [4] the behaviour of a greedy algorithm was studied proving that the greedy achieves almost surely an approximate solution whose value is one half of the value of the optimal solution. However, using a constant density model, we deal with dense graphs. What happens if we are interested in solving the problem for sparse graphs?

Erdös and Palka posed the following problem:

Let p be a function of 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 $p(n)$, for example $p=c/n$, $c>1$ constant, a random graph contains a tree of size $b(c).n$, where $b(c)$ depends only on c.

In this paper we analyze the size of the largest induced tree for a wide family of sparse graphs using a general model for random graphs. In particular, applying this analysis we will be able to prove the conjecture of Erdös and Palka.

Finally we note that a first version of the solution of the Erdös and Palka conjecture was presented at the X C.A.A.P. Conference [7].

2. A PROBABILISTIC MODEL

As we said in the introduction, in order to solve the conjecture, we would only need to prove that for a single function $p(n)$ the random graph contains a tree of size $b(c).n$. However our techniques allow to show such a result for a family of functions $p(n)$; this fact is possible using a general probabilistic model, that was already used in [5] and [6] to perform a probabilistic analysis of the max independent set problem. So first of all, we introduce this model which is a generalization of the classical models.

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

Of course the type of random graphs that we achieve with this definition depends on the value of $c(n)$. In [6] it has been shown that if the range of $c(n)$ is chosen in the real interval $(1, n)$, we start dealing with dense graph and we arrive to deal with very sparse graphs. In particular we note that c can be also a constant.

In the following, since we are essentially interested in dealing with sparse graphs we will limit the possible values of c. To simplify the notation we will write c instead of $c(n)$.

We now state two lemmas that will be used in the proofs of the next paragraph.

LEMMA 1:

$$\lim_{n \to \infty} \left(1 - c^{-c/n} - \frac{c \ln c}{n}\right) = 0.$$

Proof: The proof is trivial using some simple analytic steps.

The second lemma is more interesting. We give an upperbound, in a specific case, to the binomial which is more precise than the classical approximations.

LEMMA 2: *Let* θ *be a quantity between* 0 *and* 1. *Then*

$$\binom{n}{\theta n} \leqq (\theta^{\theta}(1-\theta)^{(1-\theta)})^{-n}.$$

Proof: The proof has been given independently in [2] and in [6].

3. MAIN RESULTS

In this paragraph we want to analyze the size of the largest induced tree for a wide family of random graphs. Instead of achieving the precise size, we will be able to give an upper and lower bound to this size. However, the two bounds are very near. In fact we will prove that, for almost every graph, there exist A and B such that the size Z_n of the largest induced tree of a sparse random graph verifies the following inequality:

$$A . n \leqq Z_n \leqq B . n, \qquad 0 < A, \quad B < 1$$

Furthermore, we will be able to give a precise analytic evaluation of A and B. First of all we state the upper bound. In the following the logarithms are to base e.

THEOREM 1. *If $p(n)$ tends to zero and c is a constant sufficiently large, then*

$$Z_n \leqq \frac{lnc + lnlnc + 1}{c/2} . n \text{ almost surely.}$$

Proof: Given a tree T, let $|T|$ denote the size of T. Let $k = \theta n$. The theorem is proved if we show that Prob $(\exists T / |T| \geqq k) \to 0$ with

$$\theta = \frac{lnc + lnlnc + 1}{c/2}$$

Let x_k be the random variable that denotes the number of induced trees of size k.

$$\text{Prob}(\exists T / |T| \geqq k) \leqq E(X_k) = \binom{n}{k} . k^{k-2} . p^{k-1} . q^{\binom{k}{2} - (k-1)} \qquad \text{where} \quad q = 1 - p$$

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

$$\leqq \binom{n}{\theta n} . k^{k-2} . (1 - c^{-c/n})^{k-1} . (c^{-c/n})^{((k-1)(k-2))/2}$$

(Exploiting Lemmas 1 and 2)

$$\leqq (\theta^\theta (1-\theta)^{(1-\theta)})^{-n} . (\theta n)^{\theta n - 2} \left(\frac{clnc}{n}\right)^{\theta n - 1} . (c^{-c/n})^{((\theta n - 1)(\theta n - 2))/2}$$

$$\left(\text{assuming } \theta \geqq \frac{lnc + lnlc + 1}{c/2}, \frac{1}{\theta^2 n} (1-\theta)^{1 - 1/\theta} c^{(c\theta - \theta n + 1)} < 1\right)$$

$$\leqq ((\theta (1-\theta)^{1/\theta - 1})^{-1} . \theta clnc . c^{-\theta c/2})^{\theta n - 1} = \left(\frac{c^{(1 - (\theta c/2))} lnc}{(1-\theta)^{1/\theta - 1}}\right)^{\theta n - 1}.$$

Now it is sufficient to consider the quantity

$$\frac{c^{(1 - (\theta c/2))} lnc}{(1-\theta)^{(1/\theta - 1)}}.$$

It is easy to see that this quantity is less than 1 giving the requested convergence to zero if

$$\theta \geqq \frac{lnc + lnlnc + 1}{c/2}$$

Q.E.D.

Before proving the lower bound we want to note that the hypotheses of the theorem are satisfied by a large class of functions $p(n)$. Equivalently this means that the result holds for every sparse random graphs and even in the case of particular dense graphs. The same considerations remain true for the lower bound.

In order to evaluate the lower bound we need two lemmas that will be crucial in the proof.

The first lemma allows to study the following situation. Fixed an induced tree T in a graph, let us consider the set I of the possible other induced trees which have l common nodes with T. We can divide I in subsets $I_0, I_1, \ldots, I_i$ according to the fact that the induced trees have $0, 1, \ldots, i, \ldots$ edges in common with T. We want to prove that the maximum cardinality of the sets I_i's is achieved for I_0, that is when we have 0 common edges.

LEMMA 3: *Let $T=(V_T, E_T)$ be a fixed induced tree of k nodes in a graph of n nodes. Let $\operatorname{Tr}(h, l)$ denote the set of the induced trees $T_{h,l}$ of k nodes having l nodes and h edges in common with T.*

Then

$$|\operatorname{Tr}(h, l)| \leqq \frac{|\operatorname{Tr}(0, l)|}{(k-l)^h}.$$

Proof: Let $L=\{1, 2, \ldots, l\}$ be the labelling of the common nodes. Let $B_{j,l}=(L, E_B)$ be the graph with $E_B \subseteqq \{(l_i, l_j) \in L/(l_i, l_j) \in E_T\}$ and $|E_B|=j$.

By definition of induced tree this implies that $B_{h,l} \subseteqq T_{h,l}$.

Now we prove that

$|\operatorname{Tr}(h-1, l)| \geqq (k-l)|\operatorname{Tr}(h, l)|$. We consider a particular $\bar{T}_{h,l}$.

Let us delete an edge $e \in B_{h,l}$. So we obtain a graph $B_{h-1,l}$ with $h-1$ edges. We can build a set $A \subseteqq \operatorname{Tr}(h-1, l)$. Every tree in A contains $B_{h-1,l}$ and is built from $T_{h,l}-\{e\}$ adding an edge $\bar{e}=(v_i, v_j)$ (with v_i or v_j but not both belonging to L) in such a way that $T_{h,l} \cup \{\bar{e}\}$ has a cycle containing e and $\bar{e}$.

Choosing every time a different edge $\bar{e}$ we build different trees in the set $\operatorname{Tr}(h-1, l)$. In this way it is easy to see that we can build at least (k-l) trees in $\operatorname{Tr}(h-1, l)$.

On the other hand applying this construction to different $T_{h,l}$ we obtain correspondingly different $(k-l)$ trees in $\operatorname{Tr}(h-1, l)$. In fact, since the edge e belongs to every $T_{h,l}$, deleting e, we obtain $|\operatorname{Tr}(h, l)|$ different graphs.

Furthermore, even in the adding phase we obtain different trees from different $\bar{e}$'s.

From the formula

$$|\mathrm{Tr}(h-1, l)| \geqq (k-l)\,|\mathrm{Tr}(h, l)|$$

the thesis follows by induction.

Q.E.D.

In the next lemma we want to find an upper bound to $|\mathrm{Tr}(0, l)|$.

LEMMA 4:

$$|\mathrm{Tr}(0, l)| \leqq \binom{k}{l} k^{k-l}(k-l)^l.$$

Proof: Cayley's Theorem gives a bijection between the set of the trees of k nodes and the set of the strings of length $k-2$ defined over the alphabet $\{1, 2, \ldots, k\}$. In particular this means that instead of considering the trees $T_{0,\,l}$ we can study the corresponding strings in order to evaluate $|\mathrm{Tr}(0, l)|$. A string that represents a tree $T_{0,\,l}$ has to verify the following property:

(1) at least $l-1$ positions in the string correspond to edges having one of the two vertices belonging to the independent set contained in $T_{0,\,l}$. Therefore in these positions we cannot have more than $k-l$ values;

(2) in the remaining positions k values are possible.

Finally we have $\binom{k}{l}$ different ways of choosing the positions corresponding to the edges of the independent set.

Therefore an upper bound to $|T_{0,\,l}|$ is given by

$$\binom{k}{l} k^{k-l}(k-l)^l$$

Q.E.D.

THEOREM 2: *If $p(n)$ tends to zero and c is a constant sufficiently large then*

$$\frac{2\,lnc-6}{clnc}.\,n \leqq Z_n \text{ almost surely.}$$

Proof: The proof is based on the second moment method and on the preceding lemmas.

As in the proof of Theorem 1 let X_k be the random variable that denotes the number of induced trees of size k. Let σ_k^2 be the variance of X_k and $E(X_k)$ the expectation of X_k. Then

$$\operatorname{Prob}(X_k=0) \leqq \frac{\sigma_k^2}{[E(X_k)]^2} = \frac{E(X_k^2)}{[E(X_k)]^2} - 1.$$

From the proof of Theorem 1 we know the value of $E(X_k)$. Therefore we need now to evaluate $E(X_k^2)$. $E(X_k^2)$ is equal to the sum of the probabilities of having ordered pairs of trees of size k with l common nodes, $1 \leqq l \leqq k$.

First of all we evaluate something slightly different, that is, $E(X_{k,h}^2)$ where $E(X_{k,h}^2)$ is equal to the sum of the probabilities of having ordered pairs of trees of size k with l common nodes *and* h common edges.

$$E(X_{k,h}^2) \leqq$$

$$\sum_{l=1}^{k} \underbrace{\binom{n}{k} k^{k-2} p^{k-1} q^{\binom{k}{2}-k+1}}_{A} \underbrace{\binom{n-k}{k-l} \cdot \binom{k}{l} |\operatorname{Tr}(h, l)| \, p^{k-h-1} q^{\binom{k}{2}-\binom{l}{2}+h-k+1}}_{B}$$

We note that respectively $A(B)$ is the part of the formula which gives the probability for the first (second) tree of the pair.

By Lemmas 3 and 4 we know that

$$|\operatorname{Tr}(h, l)| \leqq \frac{|\operatorname{Tr}(0, l)|}{(k-l)^h} \leqq \binom{k}{l} k^{k-l} (k-l)^{l-h}.$$

On the other hand

$$E(X_k^2) \leqq \max_h E(X_{k,h}^2).$$

On the whole we obtain therefore

$$\frac{\sigma_k^2}{[E(X_k)]^2} \leqq$$

$$\max_h \frac{\displaystyle\sum_{l=2}^{k} \binom{n}{k} k^{k-2} p^{k-1} q^{\binom{k}{2}-k+1} \binom{n-k}{k-l} \binom{k}{l}^2 k^{k-l} (k-l)^{l-h} p^{k-h-1} q^{\binom{k}{2}-\binom{l}{2}+h-k+1}}{[E(X_k)]^2}$$

[Remembering the value of $E(X_k)$]

$$\leqq \max_{h} \frac{\sum_{l=2}^{k} \binom{n-k}{k-l} \binom{k}{l}^2 p^{-h} q^{-\binom{l}{2}+h} (k-l)^{l-h} k^{k-l}}{\binom{n}{k} k^{k-2}}$$

(remembering that as $n \to \infty p = clnc/n$ and $k = \theta n$)

$$\leqq \max_{h} \frac{\sum_{l=2}^{k} \binom{n-k}{k-l} \binom{k}{l}^2 \left(1 - \frac{l}{k}\right)^{l-k} (\theta \, clnc)^{-h} q^{-(l/2)+h} k^2}{\binom{n}{k}}$$

(assuming $\theta \leqq (2\,lnc - 6)/clnc$, $(\theta\, clnc)^{-1} < 1$, furthermore the maximum of the sum is achieved for $h = 0$)

$$\leqq \frac{\sum_{l=2}^{k} \binom{n-k}{k-l} 4^k q^{-(l/2)} k^2}{\binom{n}{k}}.$$

(since $q^{-(l/2)} < c^{c\,\theta\, l/2}$ and the biggest term of the sum is given in the case $l = k$)

$$\leqq 4^{\theta n} c^{c\,\theta^2 n/2} \theta^{\theta n} . \theta n$$

$$\leqq (4 c^{c\,\theta/2} \theta)^{\theta n} . \theta n.$$

When p tends to zero as n tends to infinity, the last formula converges to zero exponentially if $4 c^{c\,\theta/2} \theta < 1$. This last inequality is verified if

$$\theta \leqq \frac{2\,lnc - 6}{clnc}.$$

Q.E.D.

Putting together Theorems 1 and 2 we have solved the Erdös and Palka conjecture. In fact, we have proved that there exists an induced tree of size $b(c).n$, where $b(c)$ does not depend on n but only depends on c.

Since we have an analytic expression of the lower and upper bound we can also give precise numerical approximations of the size of the largest induced tree

COROLLARY 1: *There exists a probability* $p(n)$ *such that*

$$0{,}42\,.\,n \leqq Z_n \leqq 0{,}91\,.\,n \text{ almost surely.}$$

These two bounds have been reached starting from the proofs of Theorems 1 and 2 and using some numerical approximation techniques.

Finally we note that we could have made precisely coincide the lower and upper bounds but the rate of convergence in the proofs becomes linear instead exponential. Therefore in this case the results become stronger from a mathematical point of view but less interesting from an algorithmic point of view.

Note added in proof: The conjecture by Erdös and Palka has been proved, independently by 1) Frieze, 2) Kucera and Rodl, 3) Fernandez de La Vega. The result is proved by these authors using a different approach.

REFERENCES

1. P. ERDÖS and Z. PALKA, *Trees in Random Graphs*, Discr. Math., Vol. 46, 1983.
2. 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.
3. M. KARONSKI and Z. PALKA, *On the Size of a Maximal Induced Tree in a Random Graph*, Math. Slovaca, Vol. 30, 1980.
4. A. MARCHETTI-SPACCAMELA and M. PROTASI, *The Largest Tree in a Random Graph*, Theor. Comp. Sci., Vol. 23, 1983.
5. M. PROTASI and M. TALAMO, *A New Probabilistic Model for the Study of Algorithmic Properties of Random Graph Problems*, Proc. Conf. on Foundations of Computation Theory, Borgholm, Lect. Notes in Comp. Sci., Vol. 158, 1983.
6. M. PROTASI and M. TALAMO, *A General Analysis of the Max-Independent Set and Related Problems on Random Graphs*, Tech. Rep. 3/84, Dip. Matematica, Università dell'Aquila, 1984.
7. M. PROTASI and M. TALAMO, *On the Maximum Size of Random Trees*, Proc. X Coll. on Trees in Algebra and Programming, Berlin, Lect. Notes in Comp. Sci., Vol. 185, 1985.