# Combinatorial optimization problems for which almost every algorithm is asymptotically optimal

Wojciec Szpankowski

UNITÉ DE RECHERCHE
INRIA-ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P. 105
78153 Le Chesnay Cedex
France
Tél.:(1) 39 63 55 11

Rapports de Recherche

1992

25ème anniversaire

N° 1770

*Programme 2*
*Calcul Symbolique, Programmation*
*et Génie logiciel*

# COMBINATORIAL OPTIMIZATION PROBLEMS FOR WHICH ALMOST EVERY ALGORITHM IS ASYMPTOTICALLY OPTIMAL

Wojciech SZPANKOWSKI

Octobre 1992

*RR.1770*

# COMBINATORIAL OPTIMIZATION PROBLEMS FOR WHICH ALMOST EVERY ALGORITHM IS ASYMPTOTICALLY OPTIMAL!

October 6, 1992


Wojciech Szpankowski*
Department of Computer Science
Purdue University
W. Lafayette, IN 47907
U.S.A.



## Abstract

Let $Z_{\max}$ and $Z_{\min}$ be respectively the maximum and minimum of the objective function in a combinatorial problem for which the cardinality of the set of feasible solutions is $m$ and the size of every feasible solution is $N$. We prove that in a certain probabilistic framework $Z_{\max} \sim Z_{\min}$ almost surely (a.s.) *provided* $\log m = o(N)$ for $N$ and $m$ become large. This result implies that for such a class of combinatorial optimization problems almost *every* algorithm finds asymptotically optimal solution! The quadratic assignment problem, the location problem on graphs, and some pattern matching problems fall into this class.


# PROBLEMES COMBINATOIRES POUR LESQUELS PRATIQUEMENT TOUS LES ALGORITHMES SONT ASYMPTOTIQUEMENT OPTIMAUX


## Résumé

Nous notons respectivement $Z_{\max}$ et $Z_{\min}$ les plus grandes et plus petites valeurs d'une fonction d'objectif apparaissant dans un problème de combinatoire pour lequel le nombre de solutions possibles est $m$ et la "taille" de chacune des solutions est $N$. Nous prouvons sous certaines hypothèses probabilistes que $Z_{\max} \sim Z_{\min}$ presque sûrement, du moment que $\log m = o(N)$ pour $N$ et $m$ grands. Cette propriété a pour conséquence que pour chacun de ces problèmes d'optimisation combinatoire de cette classe, presque tous les algorithmes de recherche trouve une solution optimale d'un point de vue asymptotique. Comme éléments de cette classe on peut citer : le problèmes de l'assignement quadratique, le problème de la localisation dans les graphes, et quelques problèmes de recherche de motifs dans les textes.



*This research was primary done while the author was visiting INRIA, Rocquencourt, France, and he wishes to thank INRIA (projects ALGO, MEVAL and REFLECS) for a generous support. In addition, support was provided by NSF Grants CCR-9201078, NCR-9206315 and INT-8912631, by Grant AFOSR-90-0107, and in part by NATO Collaborative Grant 0057/89.

## 1. INTRODUCTION

We consider in this paper a class of optimization problems that can be formulated as follows: for some integer $n$ define $Z_{\max} = \max_{\alpha \in \mathcal{B}_n}\{\sum_{i \in \mathcal{S}_n(\alpha)} w_i(\alpha)\}$ ($Z_{\min}$ respectively), where $\mathcal{B}_n$ is the set of all feasible solutions, $\mathcal{S}_n(\alpha)$ is the set of all objects belonging to the $\alpha$-th feasible solution, and $w_i(\alpha)$ is the weight assigned to the $i$-th object in the $\alpha$-th solution. For example, in the traveling salesman problem [13], $\mathcal{B}_n$ represents the set of all Hamiltonian paths, $\mathcal{S}_n(\alpha)$ is the set of edges belonging to the $\alpha$-th Hamiltonian path, and $w_i(\alpha)$ is the length (weight) of the $i$-th edge. Some other examples include: the assignment problem [8], [23], the quadratic assignment problem [9], [17, 18], the minimum spanning tree [5], the minimum weighted $k$-clique problem [5], [15], geometric location problems [16], and some others not directly related to optimization such as the height and depth of digital trees [12], [20], the maximum queue length [19], hashing with lazy deletion [1], pattern matching [3], edit distance [14], [22], and so forth. We analyze this class of problems in a probabilistic framework which assumes that the weights $w_i(\alpha)$ are random variables drawn from a common distribution function $F(\cdot)$. We also assume that the cardinality of the feasible set is $m$ (i.e., $|\mathcal{B}_n| = m$) and the cardinality of $\mathcal{S}_n(\alpha)$ is $N$ for every $\alpha \in \mathcal{B}_n$.

Our interest lies in identifying a class of combinatorial problems for which $Z_{\min} \sim Z_{\max}$ (a.s.) for $N, m \to \infty$. This will imply that every solution of such an optimization problem is asymptotically optimal in the sense that the relative error $(Z_{\max} - Z_{\min})/Z_{\max}$ converges to zero in a probabilistic sense. As a simple consequence, one can pick any algorithm to solve these problems, and with high probability it will be asymptotically optimal!

More precisely, we prove that $Z_{\max} = N\mu + o(N)$ (a.s.) and $Z_{\min} = N\mu - o(N)$ (a.s.) *provided* $\log m = o(N)$ where $\mu$ is the average value of weights $w_i(\alpha)$. Hence, $\lim_{n\to\infty} \Pr\{Z_{\max} - Z_{\min} \le o(1)Z_{\min}\} = 1$, that is, $Z_{\max} \sim Z_{\min}$ (a.s.).

There are many combinatorial problems that falls under our model. We mention here the quadratic assignment problem, a class of location problems, the pattern matching problem, and so forth. We shall discuss some details of these problems in the next section.

To the best of our knowledge, the formulation of the problem and its solution is new, even if the analysis present in this paper is quite simple. There are some scattered results in this direction (cf. [3], [9], [21]), but none of them addresses this issue in its generality. There is, of course, a huge volume of literature on combinatorial optimization problems (cf. [13]) but usually one assumes $\log m = O(N)$ and every problem is treated case by case.

## 2. RESULTS

Let $n$ be an integer (e.g., number of vertices in a graph, size of a matrix, number of keys in a digital tree, etc.), and $S$ a set of objects (e.g., set of vertices, elements of a matrix, keys, etc). We shall investigate the asymptotic behaviour of the optimal values $Z_{\max}$ and $Z_{\min}$ defined as follows

$$Z_{\max} = \max_{\alpha \in \mathcal{B}_n} \{ \sum_{i \in \mathcal{S}_n(\alpha)} w_i(\alpha) \} \qquad Z_{\min} = \min_{\alpha \in \mathcal{B}_n} \{ \sum_{i \in \mathcal{S}_n(\alpha)} w_i(\alpha) \} \ , \tag{1}$$

where $\mathcal{B}_n$ is a set of all feasible solutions, $\mathcal{S}_n(\alpha)$ is a set of objects from $\mathcal{S}$ belonging to the $\alpha$-th feasible solution, and $w_i(\alpha)$ is the weight assigned to the $i$-th object in the $\alpha$-th feasible solution. Throughout this paper, we adopt the following assumptions:

(A) The cardinality $|\mathcal{B}_n|$ of $\mathcal{B}_n$ is fixed and equal to $m$. The cardinality $|\mathcal{S}_n(\alpha)|$ of the set $\mathcal{S}_n(\alpha)$ does *not* depend on $\alpha \in \mathcal{B}_n$ and for all $\alpha$ it is equal to $N$, i.e., $|\mathcal{S}_n(\alpha)| = N$.

(B) For all $\alpha \in \mathcal{B}_n$ and $i \in \mathcal{S}_n(\alpha)$ the weights $w_i(\alpha)$ are identically and independently distributed (i.i.d.) random variables with common distribution function $F(\cdot)$, and the mean value $\mu$, the variance $\sigma^2$, and the third moment $\mu_3$ are finite.

Assumption (B) defines a *probabilistic model* of our problem (1). In our main result below, assumption (B) can be boldly relaxed by imposing only stationarity and some mixing conditions on the weights. Also, extensions of our assumption (A) are possible. We shall not explore these extensions in the paper.

Our main result can be summarized as follows.

**Theorem.** *The following holds as $N, m \to \infty$*

$$Z_{\min} = N\mu - o(N) \qquad \text{(a.s.)} \qquad Z_{\max} = N\mu + o(N) \tag{2}$$

*provided*

$$\log m = o(N) \ . \tag{3}$$

**Proof.** We first prove (2) only for the convergence in probability, and then extend it to almost sure convergence. Below, we consider only $Z_{\max}$. The lower bound trivially follows from the *Ergodic Theorem* (cf. [4]) and the fact that $\max_{\alpha \in \mathcal{B}_n}\{\sum_{i \in \mathcal{S}_n(\alpha)} w_i(\alpha)\} \geq E\{\sum_{i \in \mathcal{S}_n(\alpha)} w_i(\alpha)\} = N\mu$. We focus now on the upper bound.

Note that we can rewrite (1) as

$$Z_{\max} = N\mu + \sigma\sqrt{N} \max_{\alpha \in \mathcal{B}_n} \left\{ \frac{\sum_{i \in \mathcal{S}_n(\alpha)} w_i(\alpha) - N\mu}{\sigma\sqrt{N}} \right\} \ . \tag{4}$$

Let $X_\alpha = (\sum_{i \in \mathcal{S}_n(\alpha)} w_i(\alpha) - N\mu)/\sigma\sqrt{N}$. Then, our optimization problem is equivalent to $Z'_{\max} = \max_{\alpha \in \mathcal{B}_n}\{X_\alpha\}$.

Let $F_N(x) = \Pr\{X_\alpha \le x\}$. From Feller [7] (Chap. XVI.7) we know that

$$\frac{1 - F_N(x)}{1 - \Phi(x)} = (1 - o(1)) \exp(\lambda_1 x^3/\sqrt{N}) \qquad \text{where} \qquad \lambda_1 = \frac{\mu_3}{6\sigma^3} , \tag{5}$$

and $\Phi(x)$ is the distribution function of the standard normal distribution. Now , by (5) and Boole's inequality

$$\Pr\{\max_{\alpha \in \mathcal{B}_n} X_\alpha < x\} \le (1 + o(1)) m (1 - \Phi(x)) \exp(\lambda_1 x^3/\sqrt{N}) .$$

Define $a_m$ as the smallest solution to the following equation

$$m(1 - \Phi(a_m)) = 1 , \tag{6}$$

and observe that asymptotically $a_m \sim \sqrt{2 \log m}$ (cf. [10]). Then, the inequality in the last display becomes for any $\varepsilon > 0$

$$\Pr\{\max_{\alpha \in \mathcal{B}_n} X_\alpha < a_m + \varepsilon\} \le (1 + o(1)) m (1 - \Phi(a_m + \varepsilon)) \exp(\lambda_1 a_m^3/\sqrt{N}) .$$

But asymptotically $1 - \Phi(a_m + \varepsilon) \le (1 - \Phi(a_m)) e^{-2\varepsilon a_m^2}$, and together with (6), this implies

$$\Pr\{\max_{\alpha \in \mathcal{B}_n} X_\alpha < a_m + \varepsilon\} \le (1 + o(1)) \exp(-a_m^2 (2\varepsilon - \lambda_1 a_m/\sqrt{N}) .$$

Finally, as long as $a_m/\sqrt{N} = o(1)$ (cf. (3)) one can find such $\delta > 0$ that

$$\Pr\{\max_{\alpha \in \mathcal{B}_n} X_\alpha < a_m + \varepsilon\} \le \frac{1}{m^\delta} \tag{7}$$

which completes the proof of (2) for the convergence in probability.

To prove the stronger almost sure convergence result, we need some additional considerations. Note that (7) does not yet warrant an application of the Borel-Cantelli Lemma. Let $Z_m = \max_{1 \le m}\{X_i\}$, and observe that $Z_m$ is a nondecreasing sequence such that $Z_m \sim \sqrt{2 \log m}$ (pr.) with the rate of convergence as in (7). Fix now $s$, and find such $r$ that $s2^r \le m \le (s+1)2^r$. The subsequence $Z_{s2^r}$ almost surely converges to $\sqrt{2 \log s2^r}$ by the Borel-Canetlli Lemma. Due to monotonicity of $Z_m$ we also have for any $m$

$$\limsup_{m \to \infty} \frac{Z_m}{\sqrt{2 \log m}} \le \limsup_{r \to \infty} \frac{Z_{(s+1)2^r}}{\sqrt{2 \log (s+1) 2^r}} \cdot \frac{\sqrt{2 \log (s+1) 2^r}}{\sqrt{2 \log s 2^r}} = 1 \quad \text{(a.s.)} ,$$

and this completes the proof of the Theorem. ∎

**Remark**. In fact, from the proof one may conclude the following refinment of the upper bound: $Z_{\max} - N\mu = O(\sqrt{2\sigma^2 N \log m})$. It should be noted that the second term is of order $O(N)$ when $\log m = O(N)$, and our results brakes down. Nevertheless, even in the case $\log m = o(N)$ the second term may contribute significantly to the asymptotics, and in practice it cannot be completely ignored.

A direct consequence of our Theorem is the following corollary.

**Corollary**. *Let condition (3) holds. Then,*

$$\lim_{m\to\infty} \Pr\{Z_{\max} - Z_{\min} \le o(1)Z_{\min}\} = 1 \tag{8}$$

*provided* $N, m \to \infty$. ∎

The above corollary says that *any* algorithm of our optimization problem almost always finds a good (i.e., asymptotically optimal) solution, *provided* condition (3) holds. Below, we discuss three well known combinatorial problem that fall under our assumptions.

In passing, we note that assumption (B) can be substantially relaxed. Indeed, the lower bound holds for all weights that form a stationary ergodic sequence. For the upper bound, we need an extension of (4) which holds for some stationary sequences with appropriate mixing conditions (cf. [4]).

## 3. EXAMPLES

In this section we discuss in some details three optimization problems, namely, the quadratic assignment problem, the location problem, and the pattern matching problem.

### 3.1 The Quadratic Assignment Problem

Let $A = (a_{ij})$ and $B = (b_{ij})$ be two real $n \times n$ matrices, and let $\pi(\cdot)$ be a permutation of $\{1, 2, \ldots, n\}$. Then, the *quadratic assignment problem* (QAP) is defined as

$$Z_{\min} = \min_{\pi \in \mathcal{B}_n} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} b_{\pi(i)\pi(j)} \right\} \tag{9}$$

where $\mathcal{B}_n$ is the set of all permutations of $\{1, \ldots, n\}$. Clearly, the QAP falls into our general formulation (1) with $N = n^2$ and $m = n!$. Note that $\log m \sim n \log n = o(n^2)$, so our condition (3) holds. Therefore, if our assumption (B) is satisfied (e.g., this will hold if the matrices are generated independently from a common distribution), then our Theorem holds and $Z_{\min} \sim Z_{\max} \sim n^2\mu$ (a.s.) where $\mu = Ea_{ij}Eb_{ij}$. In fact, from the remark after the Corollary, we know that $Z_{\min} - n^2\mu = O(n^{3/2}\sqrt{\log n})$, as also proved by Rhee [18] in a more sophisticated probabilistic model. For some other references see [9], [17].

In passing, we should note that the *linear assignment problem* (LAP) does not fall into our category. In this case, as single matrix $A$ is given, and

$$Z_{\min} = \min_{\pi \in \mathcal{B}_n} \sum_{i=1}^{n} a_{i\pi(i)} \ .$$

Then, $N = n$ and $m = n!$, and hence $\log m \neq o(N)$. Our Theorem does not apply to this situation. In fact, for the uniform distribution of weights we know that $1.43 \leq EZ_{\min} \leq 2$ (cf. [6], [11]). It is conjectured that $EZ_{\min} \sim \pi^2/6 \approx 1.67\ldots$. On the other hand, it is easy to prove that for the exponential distribution of weights $Z_{\max} \sim n \log n$ (pr.) while for the normally distributed weights $Z_{\max} \sim n\sqrt{2\log n}$ (pr.) (cf. [8], [15], [21], [23]).

### 3.2 Location Problem on Graphs

A general location problem can be formulated as follows. Let $x_1, x_2, \ldots, x_n$ be a given set of points. The median problem selects $L$ points $c_1, c_2, \ldots, c_L$ so as to minimize (maximize) the distance between these points and the points $x_1, x_2, \ldots, x_n$. To formulate the problem in terms of our general optimization problem (1), we introduce a distance function (random variable) $d(x_i, x_j)$ which represents weights for a pair $(x_i, x_j)$. As a feasible solution $\alpha = (c_1, \ldots, c_L)$, we accept any choice of $L$ points out of $n$, so that cardinality of $|\mathcal{B}_n| = \binom{n}{L}$. Then, we have (cf. [16])

$$Z_{\min} = \min_{\alpha \in \mathcal{B}_n} \sum_{i=1}^{n-L} \min_{1 \leq j \leq L} \{d(x_i, c_j)\} \ .$$

Some simplification of the problem can be achieved if one considers the location problem on a (complete directed) graph. Indeed, let $w_{ij}$ be a weight assigned to the $(i,j)$-edge with the distribution function $F(\cdot)$. By a feasible solution, we understand a subset $\alpha = \{c_1, \ldots, c_L\} \subset \mathcal{M} = \{1, 2, \ldots, n\}$ of cardinality $L$ of vertices in a complete graph $K_n$. Then, the $L$ median problem becomes

$$Z_{\max} = \max_{\alpha \in \mathcal{B}_n} \sum_{i \in \mathcal{M} - \alpha} \max_{j \in \alpha} \{w_{ij}\} \ .$$

Note that $|\mathcal{B}_n| = \binom{n}{L} \sim n^L/L!$ for bounded $L$. Let us define $W_i(\alpha) = \max_{j \in \alpha} w_{ij}$. Note that under assumption (B) the distribution $F_W(x)$ of $W_i(\alpha)$ is $F^L(x)$. The average value $EW$ of $W_i(\alpha)$ is rather easy to evaluate in most interesting cases. For example, if the weights are exponentially distributed, then $EW = H_L$ where $H_L$ is the $L$-th harmonic number; if the weights are uniformly distributed on $(0,1)$, then $EW = L/(L+1)$, and so forth (cf. Galambos [10]). Since, $m = |\mathcal{B}_n| = n^L/L!$, and $N = n - L$, then for bounded $L$ our condition (3) of Theorem holds, and therefore

$$Z_{\min} \sim Z_{\max} = (n - L)EW + O(\sigma_W \sqrt{2nL \log n}) \sim (n - L)EW \ .$$

In particular $Z_{\min} \sim Z_{\max} \sim (n - L)H_L$ for the exponential distribution of weights, and $Z_{\max} \sim Z_{\min} = (n - L)L/(L + 1)$ for the uniformly distributed weights.

### 3.3 Pattern Matching Problem

We consider the following string matching problem: Given are two strings, a *text* string $\mathbf{a} = a_1a_2...a_n$ and a *pattern* string $\mathbf{b} = b_1b_2...b_k$ of lengths $n$ and $k$ respectively, such that symbols $a_i$ and $b_j$ belong to a $V$-ary alphabet $\Sigma = \{1, 2, ..., V\}$. The alphabet may be finite or not. Let $C_i$ be the number of positions at which the substring $a_ia_{i+1}...a_{i+k-1}$ agrees with the pattern $\mathbf{b}$. That is, $C_i = \sum_{j=1}^{k} equal(a_{i+j-1}, b_j)$ where $equal(x, y)$ is one if $x = y$, and zero otherwise (the index $j$ that is out of range is understood to stand for $1 + (j \bmod n)$). We are interested in the quantity

$$M_{m,k} = \max_{1\leq i\leq n}\{C_i\} = \max_{1\leq i\leq n}\{\sum_{j=1}^{k} equal(a_{i+j-1}, b_j)\}$$

which represents the best matching between $\mathbf{b}$ and any $m$-substring of $\mathbf{a}$, and could be viewed as a measure of similarity between these strings. Clearly, the above problem falls into our general formulation with $m = n$ and $N = k$.

We analyze $M_{m,k}$ under the following probabilistic assumption: *symbols from the alphabet $\Sigma$ are generated independently, and symbol $i \in \Sigma$ occurs with probability $p_i$*. This probabilistic model is known as the *Bernoulli model* [20]. It is equivalent to our assumption (B). From our Theorem we conclude that $M_{n,k} \sim kP$ (a.s.) provided $\log n = o(k)$, where $P = \sum_{i=1}^{V} p_i^2$ is the average value of a match in a *given* position. We observe that in the case $\log n = O(k)$, Arratia *et al.* proved that $M_{n,k} \sim (1/h)\log k$ where $h = -\sum_{i=1}^{V} p_i \log p_i$ is the entropy of the alphabet.

From the proof of Theorem we also conclude that for the case $\log n = o(k)$ we have $M_{n,k} \sim kP + O(\sqrt{2(P - P^2)k\log n})$ (pr.). However, a precise estimate of the second term in the above asymptotics is quite involved. Recently, Atallah *et al.* [3] proved that for a wide range of input probabilities $p_i$ the following is true: $M_{n,k} \sim kP + \sqrt{2(P - T)k\log n}$ (pr.) where $T = \sum_{i=1}^{V} p_i^3$.

## References


[1] D. Aldous, M. Hofri and W. Szpankowski, Maximum Size of a Dynamic Data Structure: Hashing With Lazy Deletion Revisited, *SIAM J. Computing*, 21, pp. 713-732, 1992.

[2] Arratia, R., Gordon, L., and Waterman, M., The Erdös-Rényi Law in Distribution, for Coin Tossing and Sequence Matching, *Annals of Statistics*, 18, 539-570, 1990.

[3] M. Atallah, P. Jacquet and W. Szpankowski, Pattern Matching with Mismatches: A Randomized Algorithm and Its Analysis, *Random Structures & Algorithms*, to appear.

[4] P. Billingsley, *Convergence of Probability Measures*, John Wiley & Sons, New York 1968.

[5] Bollobas, B., *Random Graphs* Academic Press, London (1985).

[6] M. Dyer, A. Frieze, and C.J.H. McDiarmid, On Linear Programs with Random Costs, *Mathematical Programming*, 35, 3-16, 1986.

[7] Feller, W., *An Introduction to Probability Theory and its Applications*, Vol. II, John Wiley & Sons, New York (1971).

[8] Frenk, J., van Houweninge, M. and Rinnooy Kan, A., Order Statistics and the Linear Assignment Problem, *Computing*, 39, pp. 165–174 (1987).

[9] Frenk, J., van Houweninge, M. and Rinnooy Kan, A., Asymptotic Properties of the Quadratic Assignment Problem, *Mathematics of Operations Research*, 10, 100– 116 (1985).

[10] J. Galambos, *The Asymptotic Theory of Extreme Order Statistics*, Second Edition, Robert E. Krieger Publishing Company, Malabar, Florida (1987).

[11] M. Karonski and W. Szpankowski, Linear Assignment Problem Revisited: Average $(1+1/e)$-out Degree Bipartite Graph Has a Perfect Matching, (Working Paper) Purdue University, CSD-TR-92-038, 1992,

[12] D. Knuth, *The Art of Computer Programming. Sorting and Searching.* vol. III, Addison-Wesley, Reading, Mass. (1973).

[13] Lawler, E., Lenstra, J.K., Rinnooy Kan, A.H., and Shmoys, D., *The Traveling Salesman Problem*, John Wiley & Sons, Chichester (1985).

[14] G. Louchard and W. Szpankowski, Probabilistic Analysis of a String Editing Problem, University of Bruxells, TR-217 (1992).

[15] Lueker, G., Optimization Problems on Graphs with Independent Random Edge Weights, *SIAM J. Computing*, 10, pp. 338–351 (1981).

[16] Papadimitriou, C., Worst-Case and Probabilistic Analysis of a Geometric Location Problem, *SIAM J. Computing*, 10, pp. 542–557 (1981).

[17] W.T. Rhee, A Note on Asymptotic Properties of the Quadratic Assignment Problem, *Opertaions Research Letters*, 7, 197-200 (1988).

[18] W.T. Rhee, Stochastic Analysis of the Quadratic Assignment Problem, *Mathematics of Operations Research*, 16, 223-239 (1991).

[19] S. Sadowsky and W. Szpankowski, Maximum Queue Length and Waiting Time Revisited: Multiserver G|G|c Queues, *Probability in the Engineering and Informational Science*, 6, pp. 157-170, 1992.

[20] W. Szpankowski, On the height of digital trees and related problems, *Algorithmica*, 6, pp. 256-277, 1991.

[21] W. Szpankowski, Combinatorial optimization through order statistics, *Second Annual International Symposium on Algorithms*, Taiwan, 1991.

[22] E. Ukkonen, A linear-time algorithm for finding approximate shortest common superstring, *Algorithmica*, 5, 313-323 (1990).

[23] Weide, B., Random Graphs and Graph Optimization Problems, *SIAM J. Computing*, 9, pp. 552–557 (1980).

ISSN 0249-6399