Station Location — Complexity and Approximation

Steffen Mecke¹, Anita Schöbel², Dorothea Wagner¹

University of Karlsruhe, {mecke,dwagner}@ira.uka.de
 University of Göttingen, schoebel@math.uni-goettingen.de

Abstract We consider a geometric set covering problem. In its original form it consists of adding stations to an existing geometric transportation network so that each of a given set of settlements is not too far from a station. The problem is known to be \mathcal{NP} -hard in general. However, special cases with certain properties have been shown to be efficiently solvable in theory and in practice, especially if the covering matrix has (almost) consecutive ones property. In this paper we are narrowing the gap between intractable and efficiently solvable cases of the problem. We also present an approximation algorithm for cases with almost consecutive ones property.

Keywords. Station Location, facility location, complexity approximation

1 Introduction

The Station Location problem consists of placing new stops along a given public transportation network in order to reach all potential customers. Building and maintaining new stops causes, of course, additional costs to the maintainers of the network. So the goal is to minimize the number of new stations to be built. The problem has arisen from a project with the largest German railway company with the goal of improving attractiveness of the public transportation network. A formal definition of the problem is the following:

Problem 1 (Station Location, cf. Fig. 1). Given a geometric graph G = (V, E), i.e. a set V of vertices in the plane (stations, switches, bends) and a set E of edges (tracks or lines) represented as straight line segments, a set P of points in the plane (settlements or demand points), and a fixed radius R. Find minimum set of vertices S (new stops) on the edges that cover P, i.e. $P \subseteq \text{cov}(S)$, where $\text{cov}(S) = \{x \in \mathbb{R}^2 : d(x, S) \leq R\}$. In the weighted version there are costs associated with the edges, and the goal is to minimize the sum of costs.

This \mathcal{NP} -hard problem has, in the form presented here, been brought up in [1] while similar problems have been studied already before. The authors of [2] consider a variant of station location with similar structural properties. Some of

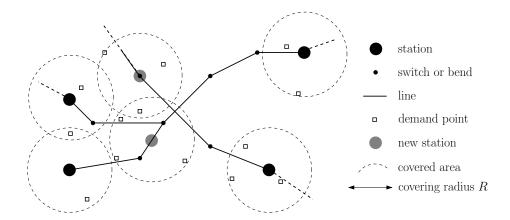


Figure 1. Problem example

the special cases mentioned there are also of interest in our case. Station Location is closely related to several other hard optimization problems, especially k-median, k-center, and facility location problems, as well as the Covering by Discs problem (see [3,4]). It can also be seen as a Set Covering problem which has many applications and has been widely studied. For an overview of work on Set Covering problem see [5].

Experimental studies have shown that most practical instances of the problem can be solved rather quickly ([6]). Many (sub-)instances of the Set Covering problem occurring in practice have the so called consecutive ones property, that is, the ones in each row of the matrix of the set covering problem occur consecutively (see Def. 2). This property is often fulfilled due to the geometric setting of Station Location. It plays an important role in explaining the relatively well-behaved nature of real-world instances of the Station Location problem. In [7] and [8] this topic is further illuminated. In this paper we try to further converge with theoretical results to the practical findings by identifying specializations of the Station Location problem that are still hard on the one hand, and giving efficient algorithms for approximating and solving quite general variants of the problem on the other hand. Note that our "positive" algorithmic results are usually valid for the more general Set Covering problem while the "negative" hardness results concern the more special Station Location.

Overview of the paper: Section 2 gives some basic definitions and summarizes the negative results, namely the \mathcal{NP} -completeness of special cases of the problem. Section 3 describes an approximation algorithm based on a block-based reformulation of the integer linear program of the SET COVERING problem. And Section 4 shortly describes an approach to get a grip on the hardness of the problem by means of parameterized complexity.

$2 \mathcal{NP}$ -Hardness Results

It was shown in [9] that there exists a finite set C of *candidate* locations for new stops which contains an optimum solution and which can be computed by an algorithm which is polynomial in the sizes of G and P. In the following we will assume that such a set of candidates is given.

Definition 1. Let $\mathcal{P} := \{1, \dots, M\}$ and $\mathcal{C} := \{1, \dots, N\}$. The matrix $A^{\text{cov}} = (a_{pc})$ with $a_{pc} := \begin{cases} 1 & \text{if } c \text{ covers } p \\ 0 & \text{otherwise} \end{cases}$ (for all $p \in \mathcal{P}, c \in \mathcal{C}$) is called the covering matrix of an instance of Station Location.

We will sometimes use the terms "demand points" and "rows" (resp. "candidates" and "columns") as synonyms.

Given the above, Station Location can be seen as a special case of the well-studied Set Covering (aka. Hitting Set) problem. We use the following notation to describe it as a linear problem:

$$\begin{array}{ll} \min & cx \\ \text{s.t.} & A^{\text{cov}}x \geq \underline{1}_M \\ & x \in \{0,1\}^N, \end{array} \tag{SCP}$$

where $\underline{1}_M \in \mathbb{R}^M$ denotes the vector consisting of M ones, $c \in \mathbb{R}^N$ contains the costs of the columns, and A^{cov} is an $M \times N$ -matrix with elements $a_{mn} \in \{0,1\}$. We may assume, without loss of generality, that all rows and columns of A^{cov} have at least one non-zero entry and that the costs c_j are positive.

The goal is to find an optimal solution x^* , i.e. a solution with minimal costs, or equivalently, an optimal set $\mathcal{C}^* \subseteq \mathcal{C} = \{1, \dots, N\}$ of columns of A^{cov} , where $\mathcal{C}^* = \{n \in \mathcal{C} : x_n^* = 1\}$.

Theorem 2 ([1]). STATION LOCATION is \mathcal{NP} -hard.

Definition 2. 1. A matrix A^{cov} over $\{0,1\}$ has the strong consecutive ones property (strong C1P) if all of its rows $m \in \{1, ..., M\}$ satisfy the following condition for all $j_1, j_2 \in \{1, ..., N\}$:

$$a_{mj_1} = 1$$
 and $a_{mj_2} = 1 \Longrightarrow a_{mj} = 1$ for all $j_1 \leq j \leq j_2$.

- 2. A matrix has the consecutive ones property (C1P) if there exists a permutation of its columns such that the resulting matrix has the strong consecutive ones property
- 3. If A_m^{cov} is a row of A^{cov} let bl_m be its number of blocks of consecutive ones.

If a matrix has the consecutive ones property the permutation of the columns making the ones appear consecutively can be found by using the algorithm of [10,11]. This algorithm can be performed in O(MN). Without loss of generality we can therefore assume that a matrix with consecutive ones property is already ordered, i.e. we assume that its ones already appear consecutively in all of its rows, i.e. $bl_m = 1$ for $m = 1, \ldots, M$. We say that a SET COVERING problem has C1P if its covering matrix A^{cov} has C1P.

4

Theorem 3 ([9]). STATION LOCATION is polynomially solvable if the covering matrix has C1P.

Proof. It can be shown that a covering matrix with C1P is totally unimodular. Now remember the well-known fact that a linear program with totally unimodular matrix has an integer optimal solution. Hence, the Set Covering problem can be solved by relaxing the integer constraints and solving the (non-integer) linear program in polynomial time.

More efficient approaches for solving SET COVERING problems with C1P can be found in [12,7]. As an example, it is easy to show that the covering matrix of a STATION LOCATION problem has C1P if no settlement can be covered by new stops on more than one line. One the negative side we have

Theorem 4. Station Location is \mathcal{NP} -hard in the strong sense (with unit costs), even for the case that no settlement can be covered from more than two lines.

Proof. By reduction from PLANAR VERTEX COVER (i.e., given a planar graph G = (V, E), find a minimum set of nodes $V' \subset V$ such that for every edge, at least one of its end nodes is in V'). In [13] it has been shown that this problem remains \mathcal{NP} -complete even for planar graphs with maximum degree 6. This can be further constrained to maximum degree 3 (which has been shown already in [14]). To this end we replace every vertex of degree 6 by the gadget of eleven vertices shown in Fig. 2. A very similar procedure works for vertices of degree 4 and 5. The resulting graph G' has $|V| + 10v_6 + 8v_5 + 6v_4$ vertices, maximum degree 3, and it is still planar $(v_6, v_5,$ and v_4 are the numbers of vertices of degree 6, 5, and 4, resp., in the original graph G). It follows almost immediately that a vertex cover of size K in G exists if and only if a vertex cover of size $K' := K + 5v_6 + 4v_5 + 3v_4$ exists in G'. We will call this restricted problem PD3VC.

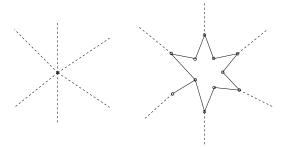


Figure 2. PLANAR DEG-6 VERTEX COVER \propto PLANAR DEG-3 VERTEX COVER

The next step is the reduction PD3VC \propto STATION LOCATION. There exists a planar orthogonal unit grid drawing of G' with $O(n^2)$ area and at most

2n+4 bends which is constructible in polynomial time (cf. [15], Theorem 4.16). We construct an instance of Station Location as follows: Let R=1/4 and construct the settlements and candidates as follows. Each edge consists of a sequence of segments of unit length in the grid. Each such segment has either two, one, or zero vertices of G' at its ends. First, replace every vertex in V by a candidate. Then, replace the segments by settlements and candidates according to the gadgets shown in Fig. 3. An example is sketched in Fig. 4.

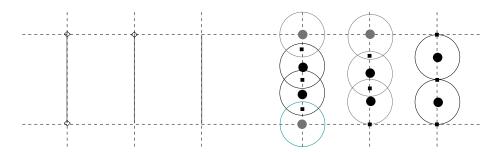


Figure 3. Three gadgets (right) for the three different types of segments (left). Settlements are depicted by squares, candidates for stations as small discs; the big circles indicate the covering radius; the grid is dashed.

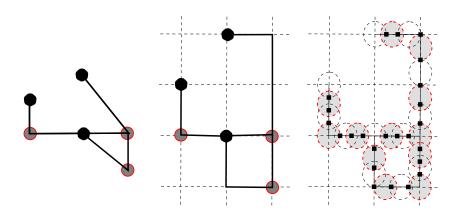


Figure 4. PD3VC \propto Station Location (with vertex cover resp. station cover in grey; candidates omitted)

Note that, after all segments have been replaced, there are exactly |V|+2|S| candidates and 3|S| settlements. Further note, that settlements are covered from candidates from different segments if and only if the corresponding segments are adjacent and no settlement is covered by more than two candidates. Finally, let

K'' := K' + |S|. A vertex cover of cardinality K' in G' exists if and only if there is a solution with cardinality K'' for the constructed instance of Station Location. As the construction works in polynomial time, this implies the \mathcal{NP} -hardness of this variant of Station Location.

Corollary 1. The (unweighted) Station Location problem remains \mathcal{NP} -complete even for the case that the covering matrix can be split into to submatrices A and B such that $A^{\text{cov}} = (A|B)$, and A and B have the consecutive ones property (even if (A|B) has exactly two 1s per row, A has no more than one 1 per row and B has no more than two 1s per row).

Proof. Consider the instance of STATION LOCATION and the graph G' constructed in the above proof. There are two classes of candidates: Candidates corresponding to vertices of G' and candidates on edges of G'. Assign columns corresponding to candidates of the first class to A and columns corresponding to candidates of the second class to B. Order the columns of B in the natural way, namely corresponding to their order on the edges between two vertices of G' (cf. Fig. 4). Then the following properties hold:

- 1. No row of A has more than one entry, because every station can only be reached by one class-A vertex. It follows that A has C1P.
- 2. An ordering of the columns of B following the above rule exists. For every row covered by columns of B the (up to two) columns covering it are consecutive.

It follows that A and B have C1P, and no row of (A|B) has more than two non-zero entries.

Note, however, the following result:

Definition 3. Let l_m (r_m) be the index of the leftmost (rightmost) 1 in the m-th row of A^{cov} . A matrix A^{cov} with strong C1P is **strictly monotone** if the sequence $(l_m)_{1 \leq m \leq M}$ and $(r_m)_{1 \leq m \leq M}$ are strictly increasing.

Lemma 1 ([7]). Let $A = (A_1|A_2)$ where A_1 and A_2 both are strictly monotone matrices. Then the SET COVERING problem with coefficient matrix A can be solved in polynomial time.

3 Approximation

It is shown in [16,17] that SET COVERING cannot be approximated within a factor of $O(\log(n))$ unless some likely assumption on complexity classes holds. Using the so called *shifting technique* of [18], however, it was proven in [19] that a PTAS exists for COVERING BY DISCS, which is similar to STATION LOCATION except for the fact that the locations for new stations are not restricted but can be chosen anywhere in the plane. The authors of [20] find a PTAS for a version of STATION LOCATION where stations cannot be arbitrarily close to each other. However, it seems that their techniques cannot be adapted to our problem (although they would probably work well in practice).

We therefore followed a different approach for approximating more general instances, especially those having only a few blocks of consecutive ones per row.

3.1 A Block-based Reformulation

Since SET COVERING problems with C1P can be solved efficiently, our idea is to split each row m with more than one block of consecutive ones (i.e. $bl_m > 1$) into bl_m rows, each of them satisfying C1P. We then require that at least one of these rows needs to be covered.

Now consider a zero-one matrix A^{cov} with M rows, such that $bl_m = 1$ for $m = 1, \ldots, p$, and $bl_m > 1$ in the remaining rows $p + 1, \ldots, M$.

For the ith block of consecutive ones in row m let

- $-f_{m,i}$ be the column of the first 1 of block i, and
- $-l_{m,i}$ be the column of its last 1.

This means that

$$a_{mn} = \begin{cases} 1 & \text{if there exists } i \in \{1, \dots, bl_m\} \text{ such that } f_{m,i} \leq n \leq l_{m,i} \\ 0 & \text{otherwise.} \end{cases}$$

Consider a row A_m^{cov} of A^{cov} with $bl_m > 1$. According to the transformation introduced in [8] we replace A_m^{cov} by bl_m rows $B_{m,1}, B_{m,2}, \ldots, B_{m,bl_m}$, each of them containing only one single block of row A_m , i.e., we define the jth element of row $B_{m,i}$ as

$$(B_{m,i})_j = \begin{cases} 1 & \text{if } f_{m,i} \le j \le l_{m,i} \\ 0 & \text{otherwise.} \end{cases}$$

Hence, due to [8], the Set Covering problem (SCP) is equivalent to

min
$$cx$$

s.t. $A_m^{\text{cov}} x \ge 1$ for $m = 1, ..., p$
 $B_{m,i} x \ge y_{m,i}$ for $m = p + 1, ..., M, i = 1, ..., bl_m$

$$\sum_{i=1}^{bl_m} y_{m,i} \ge 1$$
 for $m = p + 1, ..., M$ (SCP')

$$y_{m,i} \in \{0,1\}$$
 for $m = p + 1, ..., M, i = 1, ..., bl_m$
 $x \in \{0,1\}^N$.

It is more convenient to rewrite (SCP') in matrix form. To this end, we define

- the matrix A as the first p rows of A^{cov} ,
- $-bl = \sum_{m=p+1}^{M} bl_m$ as the total number of blocks in the "bad" rows of A^{cov} , i.e., in rows of A^{cov} without C1P,
- I as the $bl \times bl$ identity matrix,
- B as the matrix containing the bl rows $B_{m,i}$, $m = p+1, \ldots, M$, $i = 1, \ldots, bl_m$,
- C as a matrix with M-p rows and bl columns, with elements

$$c_{ij} = \begin{cases} 1 & \text{if } \sum_{m=p+1}^{p+i-1} bl_m < j \le \sum_{m=p+1}^{p+i} bl_m \\ 0 & \text{otherwise.} \end{cases}$$

(SCP') can hence be reformulated as

min
$$cx$$

s.t. Ax $\geq \underline{1}_{p}$
 $Bx - Iy \geq \underline{0}_{bl}$
 $Cy \geq \underline{1}_{M-p}$
 $x \in \{0,1\}^{N},$
 $y \in \{0,1\}^{bl}.$ (SCP")

The constraint $Cy \ge \underline{1}_{M-p}$ makes sure that at least one block of each row A_m^{cov} with $m \ge p+1$ is covered.

Note that all three matrices A, B, and C have C1P. Unfortunately, the coefficient matrix of (SCP") does not have C1P and is in general even not totally unimodular.

3.2 Approximation

From Cor. 1 we know the complexity status of SET COVERING problems with at most k blocks of consecutive ones per row: Let k be an upper bound on the number of blocks in each row of A, i.e., such that $bl_m \leq k$ for all $m = 1, \ldots, M$.

Corollary 2. For k = 1 the SET COVERING problem is polynomially solvable, for all fixed $k \geq 2$ the problem is NP-hard.

Proof. For $bl_m = 1$ the problem has the consecutive ones property and is thereby totally unimodular. For k = 2 one can use a reduction to min *vertex cover* (see [3]) to obtain a set covering problem with exactly two nonzero elements in each row; hence a set covering problem with at most two blocks per row.

To solve (SCP) we suggest Alg. 1, for which we will show that it provides a k-approximation, if k is an upper bound on the number of blocks of consecutive ones per row.

Note that Alg. 1 can be solved by linear programming, since in line 5, the coefficient matrix has C1P.

Theorem 5. Algorithm 1 is a k-approximation algorithm, where

$$k = \max_{m=1,\dots,M} bl_m.$$

Proof. Let (x^*, y^*) be an optimal solution, and (x', y') be an optimal solution of the linear programming relaxation of (SCP"). This means that

$$cx' \le cx^*. \tag{1}$$

Now note that

$$ky'_{m,i} \ge \tilde{y}_{m,i}. (2)$$

$\overline{\mathbf{Algorit}}$ **hm** 1: k-approximation

```
Input: M \times N matrix A
```

Output: approximate solution \tilde{x}

- 1 Solve LP-Relaxation of the reformulation (SCP") to obtain a solution (x', y');
- **2** for m := 1, ..., M do
- Find an index i(m) with $y'_{m,i(m)} \ge y'_{m,i}$ for all $i = 1, ..., bl_m$;
- 4 Define

$$\tilde{y}_{m,i} = \begin{cases} 1 & \text{if } i = i(m) \\ 0 & \text{otherwise.} \end{cases}$$

- **5** Solve min $\{cx : Bx \ge \tilde{y}, x \in \{0,1\}^N\}$ to obtain \tilde{x} ;
- 6 return \tilde{x} ;

This trivially holds for $\tilde{y}_{m,i} = 0$, while for $\tilde{y}_{m,i} = 1$ we know that

$$y'_{m,i} = \max_{k=1,\dots,bl_m} y'_{m,k}$$

$$\geq \frac{1}{bl_m} \sum_{k=1,\dots,bl_m} y'_{m,k}$$

$$\geq \frac{1}{bl_m} \quad \text{since } Cy \geq \underline{1}_{M-p}$$

$$\geq \frac{1}{k}$$

Moreover, $\min\{cx \colon Bx \geq \tilde{y}\} = \min\{cx \colon Bx \geq \tilde{y}, x \in \{0,1\}^N\}$, since in any optimal solution of the latter, $x \leq 1$, and the integrality constraint $x \in \mathbb{N}^N$ can be deleted since B has C1P and hence is totally unimodular. Now estimate B(kx') as

$$B(kx') = kBx' \ge ky' \ge \tilde{y},$$

where the last inequality is due to (2). In other words, kx' is feasible for $\{x \colon Bx \ge \tilde{y}\}$, and hence we get

$$kcx' \ge \min\{cx \colon Bx \ge \tilde{y}\}$$

= \pmin\{cx \cdot Bx \ge \tilde{y}, \quad x \in \{0, 1\}^N\} = c\tilde{x}

Combining the latter with (1) we finally obtain $c\tilde{x} \leq kcx' \leq kcx^*$.

4 Further Issues

Parameterized Complexity. A further means of tackling Station Location is to apply parameterized complexity techniques. It does not make much sense in our context to take the "canonical" parameter, the number of stops of the solution. We want a parameter that is small but there are usually many stops in a solution. Instead, we choose the maximum distance k between the first and

Algorithm 2: FPT Algorithm

Input: Covering Matrix A^{cov} Output: Optimal solution OPT $\in S$ Data: S, a collection of partial solutions covering rows $1, \ldots, r$ 1 Sort columns of A^{cov} lexicographically.; 2 Initialize S with the partial solution \emptyset , covering no row of A^{cov} ;

- 3 forall rows r of A^{cov} do 4 | forall partial solutions $s \in S$ that do not cover r do 5 | forall columns g of A^{cov} that cover r do 6 | Add a solution $s \cup \{g\}$ to S;
- 7 Remove s from S;
 - Remove from S all duplicate partial solutions (covering the same set of rows) except the smallest such solution;

the last non-zero entry of the covering matrix in every column. We have found the following result which is especially useful for instances which have a fairly "linear" structure and therefore their covering matrix is almost a band diagonal matrix.

Theorem 6. Station Location is solvable in $O(\text{poly}(m,n) \cdot 2^k)$ if the distance between the first and last non-zero entry is not greater than k for every column of A^{cov} . If $k = \Omega(m)$ this leads to an exponential running time. But for small values of k it can be quite efficient.

Proof (sketch). In each iteration, after step 8, S contains the optimal solution covering rows $1, \ldots r$. The correctness follows from this property. The running time follows from the fact that after each iteration all solutions in S cover rows $1, \ldots r$, and no solution covers any row after r + k, so $|S| < 2^k$.

Outlook. We are not aware of a constant factor approximation algorithm for Station Location nor a PTAS. There is no FPTAS for Station Location since it is strongly \mathcal{NP} -complete (see [21] for a comprehensive introduction to approximation algorithms). The relatively nice behaviour of practical instances can be explained by several factors. First, the geometric nature of the problem along with the fact that most settlements can be reached by only a few lines results in covering matrices that are close to having C1P. This and reduction techniques described in [6] allow large portions of the instances to be solved efficiently, resulting in relatively small problem kernels. Secondly, the distribution of settlements allows to apply the shifting technique even if it's effectiveness has not been proven theoretically for our version of the problem.

A further approach not mentioned so far could be to use techniques applied to unit disc graphs as many hard problems are easy if restricted to unit disc graphs.

References

- Hamacher, H.W., Liebers, A., Schöbel, A., Wagner, D., Wagner, F.: Locating new stops in a railway network. In: Proc. ATMOS 2001. Volume 50(1) of Electronic Notes in Theoretical Computer Science. (2001)
- Kranakis, E., Penna, P., Schlude, K., Taylor, D.S., Widmayer, P.: Improving customer proximity to railway stations. In: Proc. CIAC 2003, Rome. Volume 2654 of LNCS. (2003) 265–276
- 3. Garey, M.R., Johnson, D.S.: Computers and Intractibility. W. H. Freeman and Company (1979)
- 4. Johnson, D.S.: The NP-completeness column: An ongoing guide. J. Algorithms ${\bf 3}$ (1982) 182–195
- Caprara, A., Fischetti, M., Toth, P.: Algorithms for the set covering problem. Annals of Operations Research 98 (2000) 353–371
- Mecke, S., Wagner, D.: Solving geometric covering problems by data reduction. In: Proc. ESA '04, Norway. Volume 3221 of LNCS. (2004) 760–771
- 7. Schöbel, A.: Set covering problems with consecutive ones property. Technical report, University of Kaiserslautern (2005)
- 8. Ruf, N., Schöbel, A.: Set covering problems with almost consecutive ones property. Discrete Optimization 1 (2004) 215–228
- 9. Schöbel, A., Hamacher, H.W., Liebers, A., Wagner, D.: The continuous stop location problem in public transportation networks. Technical Report 81, Wirtschaftsmathematik, University of Kaiserslautern (2002) Submitted to a journal.
- 10. Booth, K.S., Lueker, G.S.: Testing for the consecutive ones property, interval graphs, and graph planarity using PQ-tree algorithms. Journal of Computer and System Science 13 (1976) 335–379
- 11. Meidanis, J., Porto, O., Telles, G.P.: On the consecutive ones property. Discrete Applied Mathematics 88 (1998) 325–354
- 12. Nemhauser, G.L., Wolsey, L.A.: Integer and Combinatorial Optimization. Wiley (1988)
- Garey, M.R., Johnson, D.S., Stockmeyer, L.: Some simplified NP-complete problems. In: Proc. STOC '74, ACM Press (1974) 47–63
- 14. Uehara, R.: NP-complete problems on a 3-connected cubic planar graph and their application. Technical report, Tokyo Woman's Christian University (1996)
- 15. di Battista, G., Eades, P., Tamassia, R., Tollis, I.G.: Graph Drawing. Prentice Hall, Inc. (1999)
- Lund, C., Yannakakis, M.: On the hardness of approximating minimization problems. J. of the ACM 41 (1994) 960–981
- 17. Feige, U.: A threshold of $\ln n$ for approximating set cover. J. ACM 45 (1998) 634--652
- 18. Hochbaum, D.S., Maass, W.: Approximation schemes for covering and packing problems in image processing and VLSI. J. ACM **32** (1985) 130–136
- 19. Franceschetti, M., Cook, M., Bruck, J.: A geometric theorem for approximate disk covering algorithms. Technical report, California Institute of Technology (2001)
- Galota, M., Glaßer, C., Reith, S., Vollmer, H.: A polynomial-time approximation scheme for base station positioning in umts networks. In: Proc. DIALM '01, ACM Press (2001) 52–59
- Ausiello, G., Crescenzi, P., Gambosi, G., Kann, V., Marchetti-Spaccamela, A., Protasi, M.: Complexity and Approximation. Springer (1999)