Discrete Applied Mathematics 159 (2011) 1118–1130

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

Facility location problems: A parameterized view

Michael R. Fellows [a], Henning Fernau [b,*]

[a] Charles Darwin University, Darwin, Northern Territory 0909, Australia
[b] Universität Trier, FB IV—Abteilung Informatik, 54286 Trier, Germany

ARTICLE INFO

Article history:
Received 15 September 2008
Received in revised form 10 March 2011
Accepted 25 March 2011
Available online 21 April 2011

Keywords:
Exact algorithms
Parameterized algorithms
$\mathcal{NP}$-completeness
Facility location problems
Semistructured documents

ABSTRACT

Facility location problems have been investigated in the Operations Research literature from a variety of algorithmic perspectives, including those of approximation algorithms, heuristics, and linear programming. We introduce the study of these problems from the point of view of parameterized algorithms and complexity. Some applications of algorithms for these problems in the processing of semistructured documents and in computational biology are also described.

1. Introduction

Operations Research offers an extensive literature concerned with several formalizations and variants of *facility location problems*. Such problems model the following scenario:

> A company wants to open up a number of facilities to serve their customers. Both the opening of a facility at a specific location, and the service of a particular customer through a particular facility, incurs some cost. The goal is to minimize the overall cost of opening enough facilities to serve all the customers.

We introduce the investigation of these problems from the perspective of parameterized complexity and algorithms. Properly formalized, facility location problems can also be used to model algorithmic issues in other application areas of Operations Research, Computer Science and Computational Biology.

Most variants of the class of problems that we are concerned with are $\mathcal{NP}$-hard. Motivated by their significant applications, attempts have been made to devise useful algorithms, according to three well-known strategies for "coping with $\mathcal{NP}$-hardness":

(1) Heuristics that are guaranteed to run in reasonable time and produce a solution—but with no mathematical guarantee concerning its quality.
(2) Approximation algorithms that run in reasonable time and produce a solution guaranteed to be within some distance of optimal (usually a very wide distance).
(3) A reformulation of the problem as an integer (non-)linear programming task. LP-solvers are then applied. This approach differs from (1) in that, in a reasonable amount of time, one might not get any answer, but if a solution is produced, it is optimal. This third approach has proved to be relatively successful, in part because LP-solvers have been extensively studied both in theory and in practice.

* Corresponding author.
E-mail address: fernau@uni-trier.de (H. Fernau).

0166-218X/$ – see front matter
doi:10.1016/j.dam.2011.03.021

Our research in this paper is aligned with (3). We offer an exploration of the parameterized complexity of this class of problems. We describe an abstract form of the problem, and investigate its complexity under several natural parameterizations. We also exposit how this general form of the problem connects to other applications that are seemingly far removed from locating physical facilities such as stores.

Parameterized complexity can be viewed as a multivariate generalization of the familiar *P versus NP* framework that is essentially *one-dimensional*. In this familiar framework, complexity is assessed, in the worst case, over all inputs of size (the one measurement) n. This has proven to be far too pessimistic for real-world input distributions for many computational problems. Natural input distributions tend to have important *secondary structure* that significantly affects the practical computational complexity of the problem. In the framework of parameterized complexity, the *parameter* captures the relevant secondary structure. Background on parameterized complexity can be found below.

1.1. Definitions

We study the following problem of Facility Location and variants thereof:

Given: A bipartite graph $B = (F \uplus C, E)$, consisting of a set F of potential facility locations, a set C of customers, and an edge relation E, where $\{f, c\} \in E$ indicates that c can be served from the facility (at) f; weight functions $\omega_F : F \to \mathbb{N}_{\geq 1}$ and $\omega_E : E \to \mathbb{N}_{\geq 1}$ (both called ω if no confusion may arise) that model the costs of building facilities at various locations, and the costs of serving the customers from facilities at those locations $k \in \mathbb{N}$ representing the total budget.

Question: Is there a set $F' \subseteq F$ of facility locations and a set $E' \subseteq E$ of ways to serve customers such that:

(1) every edge in E' is incident on a vertex in F', a requirement that expresses that edges used to serve customers must come from locations chosen for the facilities, or formally, $\forall f \in F (f \in F' \iff \exists e \in E' (f \in e))$,
(2) every customer is served by some edge in E', or formally, $\forall c \in C \exists e \in E' (c \in e)$, and
(3) that the budget is observed, expressed formally, $\sum_{f \in F'} \omega_F(f) + \sum_{e \in E'} \omega_E(e) \leq k$?

The set F' represents in a solution the locations where facilities are to be opened. The set E' represents how the customers should be served from those facilities.

In the literature, the problem formulated above is mostly known as the uncapacitated discrete facility location problem. See [5] for a recent overview.

Alternatively, and sometimes more conveniently, this problem can also be formulated in terms of a matrix representation of the facility location and customer service costs.

Facility Location (matrix formulation)

Given: A matrix $M \in \mathbb{N}_{\geq 1}^{(n+1)\times m}$, indexed as $M[0 \ldots n][1 \ldots m]$ $k \in \mathbb{N}$.

Question: Is there a set $F' \subseteq \{1, \ldots, m\}$ of columns and a service function $s : \{1, \ldots, n\} \to F'$ such that $\left(\sum_{f \in F'} M[0, f]\right) + \left(\sum_{c=1}^{n} M[c, s(c)]\right) \leq k$?

In the matrix formulation, the columns play the role of the m potential facility locations and the rows represent the n customers to be served (except for row 0).

Edges that are not present in the bipartite graph formulation can be modeled in the matrix representation as edges that have a weight larger than k. The matrix $M[1 \ldots n][1 \ldots m]$ records the weights of the edges, while $M[0][1 \ldots m]$ records the weights associated with potential facility locations. Condition (2) in the bipartite graph formulation can be used to construct the service function s. The equivalence of the formulations is straightforward. In the following, we will use terminology from the two formulations interchangeably, according to convenience.

1.2. Fixed parameter tractability

$\mathcal{NP}$-hard computational problems are ubiquitous in economics. One approach to overcoming this difficulty is to devise algorithms that can solve arbitrary instances of such a problem under the restriction that a relevant secondary measurement, called the *parameter*, is small.

This concept is formalized as follows. Problem instances are elements of $\Sigma^* \times \mathbb{N}$, and an instance $I = (w, k)$ is to be decided in time $\mathcal{O}(p(|w|)f(k))$, where p is a polynomial (whose degree does not depend on the parameter k) and f is an arbitrary function. Problems that can be solved within such a time restriction are said to be *fixed parameter tractable*, or in $\mathcal{FPT}$. Equivalently, a problem is in $\mathcal{FPT}$ iff there exists a polynomial time computable self-reduction κ that maps an instance $I = (w, k)$ onto an(other) instance $I' = (w', k')$ of the same problem whose overall size is limited by a function $g(k)$, i.e., $|w'| + k' \leq g(k)$. Then, I' is also called a *problem kernel* for I, and κ is the corresponding *kernelization*. There is also a complementary intractability theory, reflected in the W-hierarchy of parameterized problem classes

$$\mathcal{FPT} = W[0] \subseteq W[1] \subseteq W[2] \subseteq \ldots,$$

and an appropriate notion of parameterized problem reduction. $W[1]$-hardness corresponds to $\mathcal{NP}$-hardness in classical complexity theory. Further details can be found in the textbook [14].

The $\mathcal{O}^*$-notation extends the familiar $\mathcal{O}$-notation by suppressing factors that are polynomial in the input size n. This gives a convenient shorthand for stating results in exact exponential time and parameterized algorithmics. In particular, a