European Journal of Operational Research 262 (2017) 464–478

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Discrete Optimization

Locating names on vertices of a transaction network

David Alcaide-López-de-Pablo*, Joaquín Sicilia, Miguel Á. González-Sierra

Departamento de Matemáticas, Estadística e Investigación Operativa, Universidad de La Laguna, San Cristóbal de La Laguna, Tenerife, Spain

ARTICLE INFO

Article history:
Received 22 January 2015
Accepted 5 April 2017
Available online 20 April 2017

Keywords:
Assignment
Graph theory
Integer programming
Heuristic algorithms

ABSTRACT

This paper deals with the problem of identifying names of unknown subjects located on vertices of a network which is previously established among subjects. The only available information is the knowledge of the names of some subjects in the network, and certain records of previous observations of transactions between all pairs of subjects. Such records offer us information about the frequency or intensity of these transactions. The aim is to find the more suitable identification of the unknown subjects taking into account the information about the frequency of transactions among subjects. It is proved that the problem is NP-hard. A heuristic approach is proposed for solving it and its performance is numerically illustrated.

1. Introduction

Harary, Morgana, and Simeone (1997) studied the problem of filling the missing names of towns in a map. They formulated their problem with a graph theoretic approach. In their paper, they considered that the map showed only the names of some, but not all, towns in a region. For each town, the names of all neighbouring towns were previously known. Taking into account that information, they provided a procedure to reconstruct the missing names in the cases in which such reconstruction is possible. They also dealt with this problem for arbitrary undirected graphs, using a known list of neighbourhoods for the names.

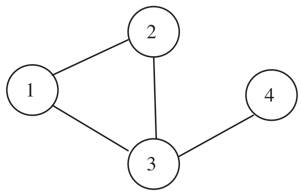
We study here a generalisation of the problem analysed by Harary et al. (1997), when the neighbours of each town are not known with complete precision. Our approach also works with arbitrary undirected graphs, but the neighbourhood among every pair of names is assessed by a value between zero and one which represents the *neighbourhood degree* between them. Even more, this point of view also allows us to study other problems in which there are relationships among subjects. These relationships can be modelled by an undirected graph, but the names of the nodes of such graph are not completely known. The problem would consist of determining, in the most suitable way, the names of all the subjects. We can refer to it as the *collaboration network problem*. In addition, we extend the undirected graph model to a directed graph one. This directed graph model represents better the situations in which the relationships are transactions between subjects and the frequency or value of such transactions is not symmetrical. We can mention it as the *transaction network problem*.

Some instances of such problems could appear in the following real situations:

(i) Let us consider a set of jobs that must be performed by a set of workers in any enterprise, industry or organisation. Each job must be performed by only one worker and any worker may perform any job. No more than one job is entrusted to each worker. For performing a pair of jobs some shared resources are necessary (for example tools, machines, computers, devices, physical space, etc.), and therefore some type of compatibility or collaboration could be required between the workers in charge of these jobs. This collaboration structure is shown by an undirected graph, where the nodes are jobs, and two nodes are adjacent if their associated jobs require share some resources. As a result, the workers assigned to such jobs must collaborate. This is the reason because we are calling this network a *collaboration network*: the jobs (vertices) which demand shared resources are linked in the network and, consequently, the workers (labels) assigned to linked jobs must collaborate among them. Consider also an affinity relationship among workers. The *affinity degree* between any pair of workers is measured by a numerical value that we can normalise to a value between zero and one. If this last value is close to one, it means that there exists better understanding and easier collaboration between the workers. We are interested to assign workers with good collaboration among them to the jobs which demand shared resources, and workers with no so easy collaboration to jobs that do not need shared resources. The problem would be to find the most suitable allocation of the workers to the jobs in such a way the total affinity/collaboration among workers assigned to linked jobs is maximised. This fact will be advantageous for the

* Corresponding author.
E-mail addresses: dalcaide@ull.es (D. Alcaide-López-de-Pablo), jsicilia@ull.es (J. Sicilia), magsierr@ull.es (M.Á. González-Sierra).

http://dx.doi.org/10.1016/j.ejor.2017.04.011
0377-2217/



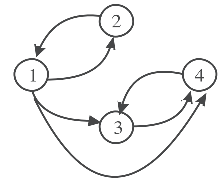
$G = (V, E)$; V = jobs = $\{1, 2, 3, 4\}$;
$E = \{(1,2), (1,3), (2,3), (3,4)\}$

$P =$

	A	B	C	D
A	–	1/2	1/3	1/2
B	1/2	–	1/2	2/3
C	1/3	1/2	–	1/2
D	1/2	2/3	1/2	–

L = workers = $\{A, B, C, D\}$

Fig. 1. Illustration of the *collaboration network problem*.



$G = (V, E)$; V = computers = $\{1, 2, 3, 4\}$;
E = transactions = $\{(1,2), (1,3), (1,4), (2,1), (3,4), (4,3)\}$

$P =$

	A	B	C	D
A	–	1/2	1/3	1/2
B	1/4	–	1/2	2/3
C	5/6	3/5	–	1/2
D	2/3	1/2	1/2	–

L = candidate names = $\{A, B, C, D\}$

Fig. 2. Illustration of the *transaction network problem*.

efficiency and productivity of the firm, and also for a more rewarding job environment among the employees.

An illustration of this collaboration network problem is shown in the particular instance drawn in Fig. 1. In this example, the jobs of the set V must be assigned to workers of the set L. At most one job will be entrusted to each worker. The links (i, j) in E represent jobs which need shared resources. The matrix $\boldsymbol{P}$ computes the *affinity degree* between workers. Therefore, the problem is to find the assignment with best total affinity degree among workers.

(ii) As second example, let us consider a set of nodes or vertices in a communication network, which is used by a set of people. These nodes could be telephone numbers in a telephone communication network, IP computer identification numbers in a computer network or keywords in the Internet, etc. Each IP or telephone number is given to only one computer or device. Someone is interested to identify the names of the people who are behind these telephone numbers, computers or devices, and who are interacting in the communication network, together with the intensity and direction of such interactions. This is the case, for instance, of the police researchers looking for contraband or delinquency networks in Internet. This transaction structure is shown by a directed graph, where the vertices are the communication network nodes (computer devices, etc.), and the arcs represent the existence of transactions between the nodes and the direction of such transactions. We will name to this framework a *transaction network*, wherein the vertices are telephone numbers, IP numbers which identify the computer devices, etc., and the arcs reflect directed transactions between them. Consequently, the people (labels) assigned to terminal vertices of each arc should perform transactions between them in the direction defined by the arc. Consider also a transaction-likelihood relationship from one person to another. The *transaction-likelihood degree* from one person to another is measured by a numerical value that we can normalise to a value between zero and one. If this last value is close to one, it means that it is highly probable the existence of *from-to* transactions between people in such direction. We are interested to assign people names to the terminal vertices of the existing arcs in the communication network with higher transaction-likelihood degree, and people names with low transaction-likelihood degree to vertices that are not linked by such arcs. The problem would be to find the most suitable allocation of the people names to the vertices of the communication network, in such a way that the total transaction-likelihood among people is maximised. This fact will be advantageous for the success of the police research because the researchers are looking for the most reliable identification of such missing names.

An illustration of this transaction network problem is shown in the particular instance represented in Fig. 2. In this figure, the computers of the network are listed in the set V. The police researchers want to break up a gang of internet delinquents. Therefore, the police should guess what people of the set L of candidates could be supposed to be behind such computers. To do it, the police know the direction of the transactions among computers. The directed arcs (i, j) in E denote the existence of communications from computer i to computer j. The matrix P collects the information about the *transaction-likelihood degree* from one person to another. Therefore, the problem is to find the most plausibility allocation of people to computers taking into account all the available information.

In this paper, we will show that these problems are harder and more difficult to solve than the Quadratic Assignment Problem (QAP). Moreover, as far as we know, they do not belong to the problems related to QAP considered in the literature. Since the first mathematical model to solve the QAP for the location of indivisible economic activities was introduced by Koopmans and Beckmann (1955), the QAP has been widely considered in the scientific literature because it can be used for modelling a great variety of real problems. Among the wide and extended literature related to the QAP, we can underline here that the QAP is known to be NP-complete (Garey & Johnson, 1979, page 218) by reduction from the Hamiltonian Circuit Problem (Sahni & Gonzalez, 1976). The QAP is even catalogued as "extremely" hard combinatorial problem for many authors as, for example, Jünger and Kaibel (2001) in their polyhedral study of the QAP-polytope. Many papers, book chapters, surveys, dissertations and books have been published through the years related to the QAP. Thus, we can cite to Burkard, Çela, Pardalos, and Pitsoulis, (1998), Burkard, Dell'Amico, and Martello (2012), Çela (1998), Hahn (1968), Karisch (1995), Lawler (1963), Loiola, Maia de Abreu, Boaventura-Netto, Hahn, and Querido (2007), Pardalos and Wolkowicz (1994), Pardalos, Rendl, and Wolkowicz (1994). A Quadratic Assignment Problem Library is published and available in Internet, which is periodically updated (Burkard, Karisch, & Rendl, 1997).

This paper is structured as follows. In Section 2 the problem is stated and a mathematical model is proposed. Section 3 analyses the computational complexity of the problem and it is proved to be NP-hard. Section 4 offers some integer programming formulations. Section 5 studies extensions of the problem. Section 6 provides a heuristic algorithm. Section 7 collects a numerical example for illustrating the approach. Section 8 shows that the problem of Harary et al. (1997) is a particular case of the problem considered in this paper. Finally, some conclusions and further research are commented in Section 9.

2. Problem statement

Suppose a *transaction network* given by a directed graph, whose set of vertices represents subjects (people, jobs, companies, telephone numbers, IP computer identification numbers, etc.) and whose set of arcs (directed) represents the existence of transactions (phone calls, e-mails, financial operations, etc.) between pairs of vertices through the direction indicated by the arc. For instance, a police research for breaking up a group of financial delinquents could require identifying a financial transaction network. The vertices are the computers used by the delinquents and the arcs represent the existence of transactions between computers in