# Planning, Designing and Evaluating Multiple eGovernment Interventions*

Fabrizio d'Amore[1], Luigi Laura[1], Luca Luciani [2] and Fabio Pagliarini[2]

[1]*Dept. of Computer, Control and Mangement Engineering, Sapienza Univ. of Rome, Via Ariosto 25, 00185 Rome, Italy*

[2]*INVITALIA Agenzia Nazionale per l'Attrazione degli Investimenti e lo Sviluppo d'Impresa S.p.A, Via Boccanelli 30, Rome, Italy*

Keywords: eGovernment, Interventions Planning.

Abstract: We consider the scenario where an organ of a public administration, which we refer as the *decisionmaker*, is requested to plan one or more interventions in some framework related to the Information Society or the eGovernment set of actions. We propose a methodology to support the decisionmaker in orienting, planning, and evaluating multiple (partially overlapping) interventions. In particular, we address two main problems: first, how to decide the structure of the interventions and how to determine the relevant parameters involved; second, how to set up a scoring system for comparing single interventions and its extension to the case of multiple interventions. The methodology unexpectedly shows that *it is not always the case that the best outcome is the one obtained by the best projects*. We formally model the problem and discuss its computational complexity. Our approach is also effective in process of selecting, from a set of submitted proposals, the ones to be funded.

## 1 INTRODUCTION

We consider a scenario aimed at planning and/or designing *interventions*, namely the definition of thematic areas, categories of users and beneficiaries, geographic locations and specific goals constituting a framework in which a *decisionmaker* wants to fund new projects, during the process of setting up an explicit call. The decisionmaker is typically a specific organ of the (central or local) public administration.

Such decisionmaker, in charge of assigning a give amount of money, has to select the type of intervention by mean of an articulated and complex decision process, which includes kind of users to benefit, type of services and level of their interactivity, state/level of existing and expected services, geographical and socio-economical context, etc. It is clear that such a decision process cannot be fully automated, but it can get benefits from the definition of guidelines and from the availability of supporting tools that make faster the so called "what-if" analysis.

Although the interventions we consider pertain the eGovernment, the Information Society and the ICT areas, the results we present may apply to several other areas.

In principle, the problem of making a decision can be modeled as a problem of optimization, defined by an objective function to be minimized (or maximized) under a set of constraints to be respected. The problem can be solved through mathematical techniques, which can be rather complex, depending on the type of objective function and constrains. In the case of multi-objective function the so-called Pareto optimum (Fudenberg and Tirole, 2002) should be searched. The number of variables and the complex constrains to be modeled might turn the resulting optimization problem intractable, that means no useful solution can be found in real cases (Garey and Johnson, 1979).

Because of such issues we here propose a novel approach, whose underlying model will be presented in the next sections. We also address the problem of evaluating interventions, by first defining the problem of evaluating a single intervention (Sect. 4), then showing how to extend it to the case of multiple interventions (Sect. 5). We formally define this problem and we prove that it belongs to the NPO complexity class[2], and therefore, if we want to efficiently solve it, we must use some heuristics.

*This work has been partially supported by the Italian National Project POSI PON ATAS.

[2]See the book of Papadimitriou (Papadimitriou, 1994) for a classical reference on computational complexity.

d'Amore F., Laura L., Luciani L. and Pagliarini F..
Planning, Designing and Evaluating Multiple eGovernment Interventions.
DOI: 10.5220/0004071300850092
In *Proceedings of the International Conference on Data Communication Networking, e-Business and Optical Communication Systems* (ICE-B-2012), pages 85-92
ISBN: 978-989-8565-23-5

Table 1: Summary of the notation used in this paper.

| *notation* | *meaning* |
|---|---|
| $B$ | available budget for all the interventions |
| $q$ | number of interventions |
| $B_i$ | budget for intervention $i$ |
| $p_i$ | number of projects to be funded in the intervention $i$ |
| $B_{i,j}$ | funding for project $j$ of intervention $i$ |
| $r_i$ | $B_{i,1}/B_{i,p_i}$ ratio between min and max funding in intervention $i$ |
| $R$ | $B_1/B_q$ ratio between the budgets of the interventions with min and max budget |

## 2 INTERVENTIONS AND BUDGET

An intervention can be characterized by: the available budget, to be granted to co-funded projects; constraints on the employment of the budget, deriving from laws and rules; types of objectives of fundable projects; category of beneficiaries and their socio-economical/territorial positions; type and impact of the expected results.

The available budget is often an amount not subjected to decision. This happens when an external organization (e.g., the European Committee) makes available to the decisionmaker an amount for co-funding projects satisfying some specific requirements. The budget defines natural constraints on the amounts to be assigned to the projects[3] and so it allows to approximately dimension the interventions.

If resources are fairly distributes, it is easy to estimate the number of projects to be funded, by defining the ratio between maximum and minimum funding. Denoting the available budget by $B$, the number of projects to be funded by $p$, the fund to be assigned to the $i$-th project by $B_i$ and the ratio between the minimum and the maximum funding by

$$r = \frac{\min_i\{B_i\}}{\max_i\{B_i\}}$$

being $0 < r \leq 1$, it is possible to exploit mathematical interpolation to dimension the amounts of the fundings. In the case of linear interpolation we have

$$B = \sum_{i=1}^{p} B_i = \frac{(\max_i\{B_i\} + \min_i\{B_i\})p}{2}$$

from which we get

$$\max_i\{B_i\} = \frac{2B}{p(1+r)}$$

[3]In the case of co-funding, the amount assigned to each project is at least the 30–35% of the budget of the whole project and therefore it determines its size.

If we re-number the projects accordingly to increasing fundings we get

$$B_j = B_1 + \frac{B_p - B_1}{p-1}(j-1)$$

for $j = 1, 2, \ldots, p$, with

$$B_1 = \frac{2rB}{p(1+r)}$$

Even if we have obtained these amounts by means of a simple and arbitrary linear interpolation, they are suitable to be the starting scheme of the decision-maker. Subsequent refinements will not cause, most likely, substantial changes of the amounts.

In some cases, the decisionmaker can program interventions by means of more than one call. Our approach still allows to determine the (base) amounts to be assigned to the projects. We introduce in a more compact form the used notation, assuming without loss of generality that both interventions and projects are numbered by increasing fundings. The linear interpolation immediately gives

$$B_{i,p_i} = \frac{2B_i}{p_i(1+r_i)}$$

Such formula requires to know $B_i$, which can be determined by an analogous procedure.

$$B_q = \frac{2B}{q(1+R)},\ B_1 = \frac{2RB}{q(1+R)}$$

The searched value is

$$B_i = B_1 + \frac{B_q - B_1}{q-1}(i-1)$$

The decisionmaker can therefore fix a few important parameters, such as $B$, $q$, $R$ and the $r_i$'s, and use them to compute the $p_i$'s and $B_{i,j}$'s. The whole process could require some iterations, but allows to quickly estimate the rough value of a few important quantities. This can be efficiently done exploiting a simple spreadsheet.

We conclude remarking the importance of recognizing the relationships existing among different interventions. In practice, if each intervention was independently planned, there would be no difference between to plan $q$ interventions and to plan $q$ times an intervention. What will make the quantum leap is identifying the dependencies existing among different types of interventions, setting up a hierarchical system that will allow to start well-coordinated and highly correlated tasks, according to a bottom-up approach aiming at privileging the construction of basic common infrastructures.

## 3 IMPACT ANALYSIS

We here introduce a methodology for carrying out the analysis of the impact of a planned intervention. It is based on the concept of *indicator*. Indicators have been introduced in statistics and are currently used in a variety of areas, among which the management control (Smith, 2009); here we use indicators for carrying out the analysis of the impact of interventions. An indicator is a mathematical function defined over a finite or infinite domain commonly defined as $D = D_1 \times D_2 \times \cdots \times D_n$, where each $D_i$ is a finite set of numbers (real, integer or natural) and $n \in \mathbb{N}$ describes the quantity of homogeneous data which we want to get concise information from. In the management control, statistical indicators are used to get concise information about some specific aspect of reality; depending on the type of analysis we are carrying on — pre-analysis, post-analysis, feasibility analysis, benchmarking etc. — many different categories of indicators can be used. In the recent literature there are several proposals providing sets of indicators, organized by category, level of aggregation, homogeneity, correlation etc. (see, e.g., (European Commission, 2010; eGEP, 2012; Ojo et al., 2005; Understand, 2006 )).

From what we discussed before, it is clear that the Indicators Set (IS) plays a critical role in the whole process of planning, designing, and evaluating interventions; the following points are therefore crucial:

1. The definition of a *correct* and *complete* Indicators Set able to model the scenario.
2. The indicators in the IS must be easily *measured* and constantly *monitored* before, during, and after the intervention. Information sources must be reliable for the whole duration of the process.
3. In order to improve the reliability, the IS should be chosen to be partially redundant, i.e. there should be some correlation between different indicators and, if possible, information sources should be chosen to obtain independently values of correlated indicators.

With distinct information sources providing the values of the indicators, it is possible on one side to have a precise picture of the real evolution of the intervention/project, on the other a variation in the correlation between related indicators might point out some errors in the measure or in the update of an indicator and, in the long run, can help in the assessment of the information sources themselves.

Given an indicators set $I = \{i_1, i_2, \ldots, i_n\}$, we define an *aggregation* (of the indicators) $A = \{A_1, A_2, \ldots, A_k\}$, where $A_i \subseteq I$ for any $i$ and $A_i \cap A_j = \emptyset$ for $i \neq j$; in other words, an aggregation is a partition of $I$, conceptually based on a high level of homogeneity. From the decisionmaker point of view, both indicators and aggregations belong to conceptual categories whose level is not sufficiently high. The decisionmaker prefers to reason about concrete objectives, directly related to benefits for citizens, enterprises, concerns, public administration etc. When defining a main topic for an intervention (e.g., the area of ICT) it is easy to define a set of (concrete) possibly interesting objectives $O = \{o_1, o_2, \ldots, o_m\}$. Once $O$ has been defined, we expect it very slowly changes as time passes, so that we can assume without loss of generality $O$ is fixed. For each item $o_i \in O$ it is possible to identify its correlations to some indicators in $I$ or, more simply, to elements in $A$.

In this way, when interested in an objective $o_i$, the decisionmaker can be easily informed about the involved indicators, related to $o_i$. It will be sufficient to make explicit all the correlations and store them into some suitable supporting system. Notice that we can consistently extend our assumption of static sets, what leads us to static correlations. Identifying elements of sets and their correlations can be done once; later, only limited maintenance will be required.

The decisionmaker is also interested in contextualizing information (according territory, socio-economics, politics etc.). We assume for simplicity one semantic coordinate of contextualization. Hence, we introduce a set of contexts $R = \{r_1, r_2, \ldots, r_\ell\}$ (e.g., the main politic units, or regions, of a given country). It is possible to introduce more sets of contexts, all of them to be considered as orthogonal. On the base of the context analysis, and of laws and rules, high priority objectives can defined, immediately identifying the involved indicators.

In order to describe all this knowledge we exploit the mathematical concept of *graph*; for basic definitions on graphs (simple graph, tree, forest, walk etc.) see for instance (Diestel, 2006). In particular we are

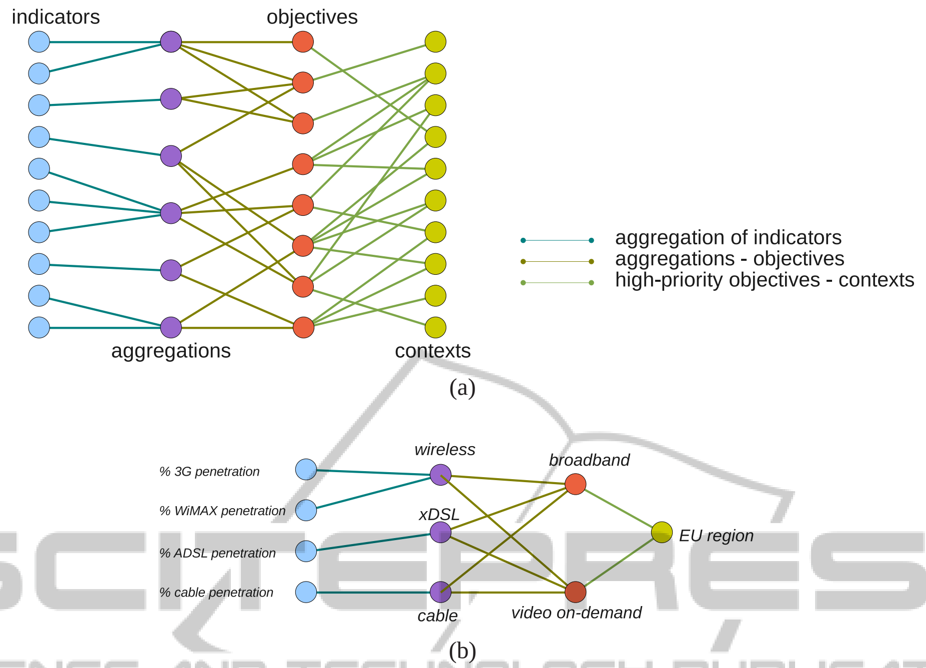




Figure 1: (a) A possible 4-parted graph, showing sets $O$, $I$, $A$ and $R$. (b) Example of tree of monotonous walks.

interested in the notion of *multipartite graph*, defined as a simple graph $G = (V, E)$ where

- $V$ is partitioned into $k$ subsets $V_i \subseteq V$, with $\bigcup_i V_i = V$ and $V_i \cap V_j = \emptyset$ for $i \neq j$;
- there is no edge $\{u, v\}$ if $u$ and $v$ belong to the same subset of vertices.

In this case the graph is said to be $k$-parted.

We can use a 4-parted graph to represent sets $I$, $A$, $O$ and $R$, and to model the correlations existing among their elements. We define a 4-parted graph whose set of vertices is defined as $I \cup A \cup O \cup R$ and it is partitioned into $I$, $A$, $O$ and $R$, and whose edges are of three types:

- Edges incident to vertices of $A$ and $I$. They model the structure of the aggregation of indicators.
- Edges incident to vertices of $O$ and $R$. They model the correlations between contexts and high-priority objectives.
- Edges incident to vertices of $A$ and $O$. They model the correlations between high-priority objectives and aggregations of indicators.

An example is given in Fig. 1 (a). Given such a graph, by selecting any vertex all related information can be automatically selected: it suffices to find the appropriate set of walks.

Given a multipartire graph, we define a *monotonous walk* as a walk having exactly one vertex in every subsets of vertices. In the 4-parted graph each monotonous walk is constituted by an indicator, an aggregation of indicators, a context and an objective. When the decisionmaker selects objective $o_i$, the set of all monotonous walks containing $o_i$ is immediately identified. It is easy to see that such a set of walks define a tree, which we call “tree of monotonous walks rooted at $o_i$.” An example is shown in Fig. 1 (b).

Our approach allows to capture the correlations among the important concepts. Notice that the model could be strengthened by quantifying the correlations, so introducing a measure that can depend not only on the two related concepts, but also on additional information (contextualization, other strongly related indicators etc.). A hypergraph (Berge, 1970), that generalizes the concept of graph, seems to be a candidate for such a quantitative model, however most of natural problems on hypergraphs are intractable. A simpler way is to use weighted graphs, by introducing a weighting function associating positive real numbers (weights) to edges.

## 4 ASSIGNING SCORES TO INTERVENTIONS

Our first need is to define a Scoring System that, at a first glance, can be seen as a block box whose input are: the target of the intervention (e.g. school, public administration, concern etc.), the location of the intervention, the state of the indicators *pre* and *post* the intervention, and the state of the average (national or international) of the values of the indicators. The output of the system is a score, representing the goodness of the intervention/project.

We now briefly describe some natural requirements that any Scoring System should satisfy; later we propose a functional scheme that meets all the requirements.

The input of the Scoring System, as described above, can be formally detailed in the following way:

- The intervention target $t \in T = \{$ set of all the possible intervention targets $\}$.
- The Indicator Domain $D = D_1 \times D_2 \times \cdots \times D_n$, where each $D_j$ is a subset of $\mathbb{R}$, the set of real numbers. Without loss of generality we can normalize all the domains to the interval $[0,1] \subset \mathbb{R}$.
- The state *pre* intervention is a vector $i_A = (i_1, i_2, \ldots i_n)$, where $i_1 \in D_1,\ i_2 \in D_2,\ \ldots, i_n \in D_n$.
- The state *post* intervention is a vector $i_P = (i_1, i_2, \ldots i_n)$, where $i_1 \in D_1,\ i_2 \in D_2,\ \ldots, i_n \in D_n$.
- The (national or international) average is a vector $i_M = (i_1, i_2, \ldots i_n)$, where $i_1 \in D_1,\ i_2 \in D_2,\ \ldots, i_n \in D_n$.
- The locality of the intervention is completely described by the vector $i_A$ (state of the indicators before the intervention).

The output of the System is a score that, without loss of generality, we can assume between 0 and 1; therefore a Scoring System can be seen as a function $f : T \times [0,1]^3 \to [0,1]$. Some natural requirements for a scoring system are:

- If $i_P = i_A$, then $f(t, i_A, i_P, i_M) = 0$ (zero score): if a project does not improve any of the indicators then its score is 0.
- If $i_P = (1, 1 \ldots, 1)$, then $f(t, I_A, i_P, i_M) = 1$ (maximum score): if an intervention/project raises all the indicator the the maximum then its score is maximum.
- Given two projects $P1$ and $P2$, with $i_{P1} = (i_1, i_2, \ldots i_j + \Delta i_j \ldots i_n)$ and $i_{P2} = (i_1, i_2, \ldots i_j \ldots i_n)$, then $f(t, i_A, i_{P1}, i_M) \geq f(t, i_A, i_{P2}, i_M)$ (non decreasing property): if two different projects bring all the indicators to the same values, except one, the project better performing on that indicator should score better (or equal[4]).

[4]The score can be equal when, given a set of weights representing the relative importance of the indicators, the corresponding weight is 0.

## 4.1 A Proposed Scoring Function

In this section we present a scoring system that satisfies the requirements described previously. In order to do so, we first need to define a way to model the intervention target, and we decide to represent it as a vector of $n$ weights $t = (w_1, w_2, \ldots, w_n)$, where $w_j \in [0,1]$ for $j = 1, 2, \ldots, n$. Here $n$ is the number of indicators and each weight $w_j$ represents the relative importance of the indicator for the given target. The vector of weights can be derived from the tree of monotonous walks previously introduced, by identifying the objective (element of set $O$) which the target is aiming at. In the case of more than objectives, the associated forest of monotonous walks will be considered.

That being stated, the scoring system can be represented by the following function:

$$f(t, i_A, i_P, i_M) = \frac{t \cdot (i_P - i_A)}{t \cdot (\overline{1}_n - i_A)} \tag{1}$$

where we have denoted by $\overline{1}_n$ the vector whose $n$ components are all equal to 1, by "$\cdot$" the vector product and by "$-$" the vector difference. We recall that, given two vectors $v = (v_1, v_2, \ldots, v_n)$ e $w = (w_1, w_2, \ldots, w_n)$ it holds that

$$v \cdot w = v_1 \cdot w_1 + v_2 \cdot w_2 + \cdots + v_n \cdot w_n$$

and

$$v - w = (v_1 - w_1, v_2 - w_2, \ldots, v_n - w_n)$$

We prove now that this function satisfies all the requirements:

- If $i_P = i_A$, then $f(t, i_A, i_P, i_M) = 0$ (zero score):

  $$f(t, i_A, i_P, i_M) = \frac{t \cdot (i_A - i_A)}{t \cdot (\overline{1}_n - i_A)} = \frac{t \cdot \overline{0}_n}{t \cdot (\overline{1}_n - i_A)} = 0$$

  where we have denoted by $\overline{0}_n$ the vector whose $n$ components are all equal to 0.
- If $i_P = (1, 1 \ldots, 1)$, then $f(t, I_A, i_P, i_M) = 1$ (maximum score):

  $$f(t, i_A, i_P, i_M) = \frac{t \cdot (\overline{1}_n - i_A)}{t \cdot (\overline{1}_n - i_A)} = 1$$
- Given two projects $P1$ and $P2$, with $i_{P1} = (i_1, i_2, \ldots i_j + \Delta i_j \ldots i_n)$ and $i_{P2} = (i_1, i_2, \ldots i_j \ldots i_n)$, then $f(t, i_A, i_{P1}, i_M) \geq f(t, i_A, i_{P2}, i_M)$ (non decreasing property):

  $$\begin{aligned} & f(t, i_A, i_{P1}, i_M) - f(t, i_A, i_{P2}, i_M) = \\ & = \frac{t \cdot (i_{P1} - i_A)}{t \cdot (\overline{1}_n - i_A)} - \frac{t \cdot (i_{P2} - i_A)}{t \cdot (\overline{1}_n - i_A)} = \\ & = \frac{t \cdot (i_{P1} - i_A) - t \cdot (i_{P2} - i_A)}{t \cdot (\overline{1}_n - i_A)} = \end{aligned}$$

$$= \frac{t\cdot(i_{P1}-i_A-i_{P2}+i_A)}{t\cdot(\overline{1}_n-i_A)} = \frac{t\cdot(i_{P1}-i_{P2})}{t\cdot(\overline{1}_n-i_A)} =$$

$$\frac{t\cdot((i_1,i_2,\dots,i_j+\Delta i_j,\dots,i_n)-(i_1,i_2,\dots,i_j,\dots,i_n))}{t\cdot(\overline{1}_n-i_A)} =$$

$$= \frac{t\cdot(0,0,\dots,\Delta i_j,\dots,0)}{t\cdot(\overline{1}_n-i_A)} \geq 0$$

We notice that this function does not keep into account the national (or international) average of the indicators; the above definitions can be easily adapted to include it.

# 5 EVALUATION OF MULTIPLE AND OVERLAPPING PROJECTS

So far we have considered the scoring of a single intervention/project. We now consider the case in which there are several distinct intervention/projects, potentially overlapping. It is important to mention that, when planning multiple projects, the value of the indicators after the projects must be carefully analysed. Let us provide an example: assume that, in a given area, the broadband penetration is 30%; we have two distinct projects, using distinct technologies, that have been estimated to raise that value by, respectively, 35% and 45%. It is clear that, when estimating the overall improvement of both projects, we cannot simply add the values, since this would lead to an unfeasible value of 110%; neither we can estimate it to 100%, because it is reasonable that there should be some overlapping in the population reached by both projects, and therefore the real value might be something slightly bigger than 75%.

Therefore it is important to analyze the effect of the multiple projects together, rather than simply summing up all the (estimated) effects. We now provide an example of a somewhat of a paradoxical effect: given a ranking of projects, it might happen that, when we want to fund some of them, the best outcome is when we choose the worst (in the ranking) projects.

Let us assume that we have 4 projects and 3 indicators; for the sake of simplicity we assume that (i) all the weights in the target vector are equal to 1 ($t = (1,1,1)$), (ii) the initial value of all the indicators is equal to 0 ($i_A = (0,0,0)$), (iii) the cost of each project is unitary, and (iv) our budget is 2, i.e. we can choose at most two projects amongst them. The post intervention vectors for the projects are as follows:

- $i_{P1} = (1.0, 0.0, 0.0)$
- $i_{P2} = (0.9, 0.0, 0.0)$
- $i_{P3} = (0.5, 0.3, 0.0)$
- $i_{P4} = (0.5, 0.0, 0.2)$

It is easy to see that, if we compute the scoring function as defined in 4, the outcome is

$$f(t, i_A, i_{P1}) > f(t, i_A, i_{P2}) > f(t, i_A, i_{P3}) > f(t, i_A, i_{P4})$$

Since there is budget for two projects, it would seem natural to fund $P1$ e $P2$; but let us now consider the post intervention vectors for all the possible pairs:

- $i_{(P1+P2)} = (1.0, 0.0, 0.0)$
- $i_{(P1+P3)} = (1.0, 0.3, 0.0)$
- $i_{(P1+P4)} = (1.0, 0.0, 0.2)$
- $i_{(P2+P3)} = (1.0, 0.3, 0.0)$
- $i_{(P2+P4)} = (1.0, 0.0, 0.2)$
- $i_{(P3+P4)} = (1.0, 0.3, 0.2)$

It is clear that, if we have to choose only two projects, the best outcome is when we fund $P3$ and $P4$, that, considered alone are worst than $P1$ and $P2$, but together are better.

We can formally define the problem above discussed (for the sake of simplicity we do not include the intervention target $t$):

MULTIPLE PROJECTS EVALUATION (MPE).
Given in input:

- an initial scenario $S$, represented by the values of a set of indicators $I = (i_1, i_2, \dots i_n)$;
- a set of projects $P = (p_1, p_2, \dots p_m)$, each associated with a cost $(c_1, c_2, \dots c_m)$ and a post intervention vector $(v_1, v_2, \dots v_m)$; with $I(p_j)$ we denote the (estimated) values of the indicators after the completion of project $p_j$; if $R \subseteq P$ with $I(R)$ we denote the (estimated) values of the indicators after the completion of all the projects in $R$.
- a scoring function $f : I \to \mathbb{R}$;
- a real number $b$, representing the available budget;

we look for a projects subset $P' \subseteq P$, whose overall cost is less than the budget, to maximize the scoring function; more formally we look for a subset $P'$ such that:

- $\sum_{j:p_j \in P'} c_j \leq b$ (budget constraint)
- $\forall P'' \subseteq P, P'' \neq P', \sum_{j:p_j \in P''} c_j \leq b,\ f(P') \geq f(P'')$ (optimality constraint)

Can we design efficient algorithms able to solve this problem? Unfortunately, the problem belongs to the NPO complexity class, as stated in the following theorem:

**Theorem 1.** *The optimization problem MPE, as defined above, belongs to the NPO complexity class.*

**Proof.** Let us now consider the decision problem associated with MPE, i.e. a the problem in which the *optimality constraint* is replaced by the following

$$f(P') \geq S$$

where $S$ is a parameter: now the problem is, given also the parameter $S$, to find a subset $P'$ such that $f(P') \geq S$. Let us denote by Decision-MPE (DMPE) this decision problem. To prove that MPE $\in$ NPO we will show that its corresponding decision problem DMPE is NP-complete (NPO is the complexity class of the optimization problems whose decision versions belong to NP (Ausiello et al., 1999)).

We recall that, in order to prove that a problem $B$ is NP-complete, it is sufficient to show what follows (see, e.g., (Garey and Johnson, 1979)):

1. There must be an NP-complete problem $A$ such that $A \leq B$ ($A$ polynomially reduces to $B$).
2. $B$ belongs to NP (e.g., by describing a polynomial time algorithm for a non-deterministic Turing Machine able to solve it).

Let us now consider the following problem (Garey and Johnson, 1979)

DOMINATING SET (DS).
INSTANCE: Graph $G = (V, E)$ and a positive integer $K \leq |V|$.
QUESTION: Is there a subset $V' \subseteq V$ such that $|V'| \leq K$, and such that every vertex $v \in V - V'$ is joined to at least one member of $V'$ by an edge in $E$?

The Dominating Set problem is NP-complete ((Garey and Johnson, 1979)); let us now define a modified version, where the underlying graph is directed:

DIRECTED DOMINATING SET (DDS).
INSTANCE: Directed graph $G = (V, E)$ and a positive integer $K \leq |V|$.
QUESTION: Is there a subset $V' \subseteq V$ such that $|V'| \leq K$, and such that every vertex $v \in V - V'$ is joined to at least one member of $V'$ by an outgoing edge in $E$?

We know prove that DMPE is NP-complete, and our proof is articulated in the following steps:

1. DS $\leq$ DDS;
2. DDS $\leq$ DMPE;
3. DMPE $\in$ NP.

**Step 1: DS $\leq$ DDS**. To show that DS reduces to DDS we simply consider the following transformation: we change every undirected edge $\{a, b\}$ of the DS instance into the two directed edges $(a, b)$ and $(b, a)$. It is easy to check that every solution of the *DDS* instance obtained in this way is also a solution of the related *DS* instance.

**Step 2: DDS $\leq$ DMPE**. To map the generic DDS instance into one DMPE instance we will:

- each node $v \in V$ is mapped into a project $p \in P$;
- each node $v$ is also mapped into an indicator $i \in I$;
- the initial value of each indicator is equal to 0;
- for each node $v \in V$, for each outgoing edge $(v, w)$ we set equal to 1 the $w$-th components of the post intervention vector associated with node $v$;
- for each node $v \in V$ we also set equal to 1 the $v$-th components of the post intervention vector associated with the node;
- each project $p_i$ is associated with a cost $c_i$=1;
- we set $b = K$;
- the objective function is $f = \sum_{\forall i \in I, i \geq 1} 1$ (it counts the number of indicators equal to 1);
- we set $S = n = |V|$.

Informally, the reduction is as follows: each node $v$ is mapped into a project able to cover all the nodes reached by $v$, together with node $v$ itself. The budget is able to select only $K$ nodes/projects and the parameter $S$ is set in such a way that all the nodes must be either dominators or dominated.

**Step 3: DMPE $\in$ NP**. To show that DMPE $\in$ NP is sufficent to observe that this problem can be solved by generating all the possible subsets of $P$ and by checking if one of them satisfies the constraint $f(P') \geq S$. A non-deterministic Turing Machine can simply *guess* at step $i$ whether to include or not the $i$-th project in the solution, and then check the constraint at the last step: therefore it takes linear time to solve it, and this imply that DMPE $\in$ NP.

## 6 CONCLUSIONS

We addressed the scenario where an organ of a public administration, i.e. the *decisionmaker*, is requested to plan one or more interventions in some framework related to the Information Society or the eGovernment set of actions.

We proposed a methodology to support the decisionmaker in orienting, planning, and evaluating multiple (partially overlapping) interventions. In particular, we address two main problems: first, how to decide the structure of the interventions and how to determine the relevant parameters involved; second, how to set up a scoring system for comparing single interventions and its extension to the case of multiple interventions. The surprising result from this formal analysis is that not always the best projects together achieve the best outcome.

We formally modeled the problem and discussed its computational complexity, showing that it is *NP*-complete the problem of the selection of the projects whose overall outcome is maximized.

## REFERENCES


Ausiello, G., Protasi, M., Marchetti-Spaccamela, A., Gambosi, G., Crescenzi, P., and Kann, V. (1999). *Complexity and Approximation: Combinatorial Optimization Problems and Their Approximability Properties*. Springer-Verlag New York, Inc.

Berge, C. (1970). *Graphes et Hypergraphes*. Dunod, Paris.

Diestel, R. (2006). *Graph Theory*. Graduate Texts in Mathematics. Springer.

eGEP (2012). eGovernment Unit - DG Information Society and Media - European Commission. *eGovernment Economics Projects - Measurements Framework*. Final Deliverable, available at http://rso.it/eGEP.

European Commission (2010). . *Key ICT indicators for the Member States, Norway and Iceland*. i2010 Annual Information Society Report 2007, Vol. 3. http://ec.europa.eu/information_society/eeurope/i2010.

Fudenberg, D. and Tirole, J. (2002). *Game Theory*. MIT Press.

Garey, M. R. and Johnson, D. S. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman and Company.

Ojo, A. K., Janowski, T., and Estevez, E. (2005). Determining progress towards e-government: What are the core indicators? In *Proc. of the 5th European Conference on e-Government*, pages 312–322. ACL, Reading, UK.

Papadimitriou, C. M. (1994). *Computational complexity*. Addison-Wesley.

Smith, C. (2009). Economic indicators. In Wankel, C., editor, *Encyclopedia of Business in Today's World*. SAGE Publications, Inc., California, USA.

Understand (2006). Understand (European regions UNDER way towards STANDard indicators for benchmarking information society). *Methodological Handbook*. Available at http://www.understand-eu.net/.