

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Bagloee, SA;Sarvi, M;Patriksson, M;Asadi, M

Title:

Optimization for Roads' Construction: Selection, Prioritization, and Scheduling

Date:

2018-10-01

Citation:

Bagloee, S. A., Sarvi, M., Patriksson, M. & Asadi, M. (2018). Optimization for Roads' Construction: Selection, Prioritization, and Scheduling. Computer Aided Civil and Infrastructure Engineering, 33 (10), pp.833-848. <https://doi.org/10.1111/mice.12370>.

Persistent Link:

<https://hdl.handle.net/11343/284439>

Optimization for roads' construction: selection, prioritization and scheduling

Saeed Asadi Bagloee , Majid Sarvi

Melbourne School of Engineering, The University of Melbourne, Victoria 3010, Australia

Michael Patriksson

Chalmers University of Technology and University of Gothenburg, SE-412 96 Gothenburg, Sweden,

&

Mohsen Asadi

Mechanical Engineering Department, University of Saskatchewan, Engineering Building, 57 Campus Drive, Saskatoon, SK Saskatoon SK S7N 5A9, Canada,

Abstract: *Limited resources (budget, labor, machinery) have a significant toll on the roads' construction. The question of interest is: given variations of resources over a lengthy construction time, what would be the best construction scheduling plan, or how to optimize the Gantt chart while considering two highly challenging features (i) prerequisite conditions and (ii) the interdependency of the benefit of the projects' completions. We formulate it as a bilevel problem where the objective function is to minimize generalized costs and the lower level accounts for the drivers' route choice. We employ a solution algorithm based on a supervised learning technique (a linear regression model of machine-learning) and an integer programming problem and it is applied to the datasets of Winnipeg and Chicago. The regression model was found to be a tight approximation which resulted in an efficient algorithm (the CPU time is almost a linear function of the number of iterations). Moreover, the proposed methodology can render promising results (at least locally optimal solutions). This paper is the first to formulate the Gantt chart using linear binary constraints and optimize it tailored to real-life case-studies.*

1. INTRODUCTION

To address chronic traffic congestion and reinvigorate aging infrastructure for future demand, investment into the road infrastructure is sometimes inevitable. On one hand, road construction is time-consuming, capital-intensive and laborious, on the other hand, capital, skilled workers, and machinery are largely scarce which calls for a knowledge-based decision support system to optimize the construction operations. The primary motive of this study stems from an intuitive and (later to be

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/mice.12370](#).

This article is protected by copyright. All rights reserved.

known) insidious request made by a client: “There is a dozen of projects approved for construction in the next five years, but our yearly budget and contractors are limited, so simply help us to prioritize the projects”. Therefore, this study seeks an answer to the following question: given the fluctuation of allocated funding and resources over the lengthy construction time, which projects should be selected and what would be the best construction plan (i.e. prioritization and the constructions’ scheduling).

To comprehend the complexity of the problem let us first have an in-depth discussion on nature and different features of the roads’ construction in five folds: (i) Due to the longevity of these projects as well as being resource intensive, a project sometimes needs to be split into several phases and some phases might be put on hold due to lack of resources which are depicted as a Gantt chart. (ii) To make it worse, some projects might be designed as prerequisites to other projects in the downstream. For example construction operations of a bridge cannot commence unless constructions of its access roads or ramps are completed, or sometimes vested interest consideration warrants such prerequisites (the prerequisite requirements pose computational challenges to other civil engineering practices, such as building constructions (Ponz-Tienda et al., 2015)). (iii) Concurrent constructions of the projects in different phases (i.e. overlapping construction phases) subject to available resources are always the case in practice, whereas, it is found computationally challenging and hence overly ignored in the literature (Bagloee and Asadi, 2015). (iv) On top of all, the complexity rests largely on the fact that benefits of the projects’ completions are tightly tied to the improvement of traffic circulation which are highly intertwined. In other words, a different order of completion of the very same projects may result in totally different benefits (not to mention, when the number of projects itself varies), which is manifested in the improvement of the traffic circulation (we call it the issue of the interdependency of benefits). To account for the improvement of traffic circulation we solve a series of traffic assignment problems (TAPs). Obviously, as the number of projects increases, the number of these combinations increases rapidly, such that, the benefit of a construction scenario cannot be assumed as a priori exogenously known input, in fact, unless it is computed endogenously. Despite the relevance and applicability of these kinds of problems ‘especially to the industry’, not much interest has been seen from the academia. Although, there exists a large body of literature on projects selection (also known as the network design problem, NDP), a combination of projects’ prioritization and constructions’ schedules are yet to be addressed. Given such a compelling relevance, this research contributes to the literature to address the roads’ selection, prioritization and (construction) scheduling problem (RSPSP) comprehensively. To this end, we formulate it as a bilevel programming problem where the objective function ‘in the upper level’ minimizes the total delay on the network as a proxy for traffic improvement (A variety of problems arising in different disciplines including civil engineering (Sarma and Adeli, 2001) can be formulated as a bilevel problem (Sinha et al., 2017)). The interconnections among the projects and phased-structure of constructions including the prerequisites requirements as well as resources’ limitations are coded using some linear integer constraints. The lower level accounts for the traffic equilibria (solving for TAPs) to compute interdependent benefits of the construction scenarios.

The RSPSP can be shown to be among the most difficult problems (in terms of computational time) namely a strong NP-hard one (Ben-Ayed and Blair, 1990). It implies that, for real-life examples (to which this paper is tailored), conventional exact methods of the optimization science, alone, are not up to this job. Therefore, as a solution method, we employ a machine-learning (ML) inspired approach in which an integer programming method is hybridized with a supervised learning technique (Bagloee et al., 2018). This methodology is tested using two real datasets, Winnipeg and Chicago.

In summary, this study contributes to the literature in two folds, (i) being the first to formulate the complicated Gantt chart, in its entirety, using some linear integer constraints, and (ii) being the first to optimize the Gantt chart for roads construction planning and scheduling tailored to real-life case-

studies while addressing the issue of the interdependency of benefits and considering practical features such as concurrent constructions, prerequisite projects, and phased-structure construction. Therefore, the two datasets used in this study are made available online to be used as a benchmark for future investigation by the research community (Bagloee 2017).

In the next section, a review of the relevant studies is presented. In Section 3, mathematical setup of the RSPSP is provided. In Section 4, the solution algorithm is introduced followed by numerical application in Section 5. We conclude the paper in Section 6.

2. LITERATURE REVIEW

For the problem of prioritization (i.e. what would be the best order or sequence of projects' completion), pioneering studies overlook the interdependency of the benefits or at best they are bound to small size examples. Nemhauser and Ullmann (1969) considered projects interdependency as only a pair-wise interaction using dynamic programming approach for a quadratic objective function. Gear and Cowie (1980) and Fox et al. (1984) discussed the extension of Nemhauser and Ullmann's model to higher-order interactions. Weingartner (1966), Cochran et al. (1971) and Johnson et al. (1985) employ an integer programming approach to the prioritization problem. Zhang et al. (2006) employ a particle swarm optimization methodology to minimizing a project duration in which no such a benefit interdependency between different activities exists while accounting for the types of resources (renewable and non-renewable). Similarly (Lim et al., 2014) take the concurrent feature of a scheduling plan to accelerate a construction project.

Jong and Schonfeld (2001) address the interdependency of projects in a sequencing problem by developing an approximation model based on simulation results. To capture the interdependency issue, simulation technique has also been used in other studies (Wang and Schonfeld, 2012). Overall, burdensome computational time of the simulation methods is a significant prohibitive factor for real-life applications.

Some studies use a traffic assignment model to evaluate projects' impacts (Tao and Schonfeld, 2005, 2006, 2007). Taplin and Qui (2001) apply a genetic algorithm and traffic assignment to the prioritization problem.

Karim and Adeli (1999b) They developed a neural dynamics model for scheduling construction with the goal of minimizing the direct construction. This model has been packaged as a patent and software (Adeli and Karim, 2003; Karim and Adeli, 1999a). The same concept has also been used for a general problem of resource scheduling (Senouci and Adeli, 2001).

Some studies managed to address the problem of prioritization for large-sized networks considering the interdependency issue via a neural network model (Bagloee and Asadi, 2015; Bagloee and Tavana, 2012). Bagloee and Asadi (2015) include also variations of the travel demand over a protracted constructions time which comes at a momentous complexity and computational cost.

In the last decade, the longevity of roads' constructions in the context of the network design problem (NDP) has been recognized (Kim et al., 2008; Lo and Szeto, 2009; Szeto et al., 2010). In fact, a series of the NDPs along construction phases is solved such that constructions' commencement time can also be considered as decision variables.

Findings of the above literature review may be summarized as follows: (i) Real-life applications is a rare currency. (ii) The interdependency of the benefits is overlooked. (iii) Given the inherent complexity of the prioritization problem, concurrent constructions of the projects is also ignored. (iv) The phased-structure of the lengthy constructions is sometimes simplified. (v) Travel demand over the construction time is typically assumed to be fixed. (vi) Prerequisite requirements in the

constructions are neglected. (vii) Only a few studies employ the TAP in their formulation to calculate the interdependent benefits of the projects' completions.

In a holistic approach, we in this study address all the above-listed shortcomings.

3. MATHEMATICAL FORMULATION

The notations used in this article is presented in Table 1. In this section, we provide a formulation for the RSPSP which is further elaborated using an illustrative example. The RSPSP may then be written as follows.

$$\text{RSPSP: } \min_{x_a^i, d_{px}^i, y_{px}^i} Z = \sum_{i=1}^n \sum_{a \in A \cup A'} x_a^i \cdot t_a(x_a^i) \quad (1)$$

subject to

$$M.(y_{px}^i - 1) + D_{px} - \sum_{j=1}^{i-1} d_{px}^j \leq 0 \quad px \in A', i = 1..n \quad (2)$$

$$\sum_{j=1}^{i-1} d_{px}^j - y_{px}^i \leq D_{px} - \varepsilon \quad px \in A', i = 1..n \quad (3)$$

$$\sum_{px \in A'} r_{px} \cdot d_{px}^i \leq R^i \quad i = 1..n \quad (4)$$

Table 1
Notations used for the RSPSP

symbol	description
A, A'	sets of existing roads and projects, respectively
px	an id label to represent a project indexed by $x = \{1, 2, \dots, A' \}$
n	total number of phases (or time-slices) of the construction
i	an index to represent a construction phase (e.g. consider the total span of construction is two years and it is divided into 8 phases or time-slices ($n = 8$) of 3 months, hence, $i \in \{1, 2, \dots, 8\}$)
R^i	available resource for the construction phase i
r_{px}	the amount of resources that project px needs during a construction phase
d_{px}^i	a main binary decision variable, with value 1, if the project px is being constructed during the phase i and 0 otherwise
D_{px}	the total number of phases that project px needs in order to complete the construction
y_{px}^i	an (auxiliary) binary decision variable; 1: if construction of the project px is already completed (in the previous phases) and it is operational during the phase i , and 0 otherwise*
x_a^i	traffic flow on the link $a \in A \cup A'$ during the construction phase i

$t_a(x_a^i)$	travel cost or time or delay of the link $a \in A \cup A'$ (pertinent to the construction phase i), defined by a non-negative and non-decreasing differentiable function derived and calibrated from a field survey.
Q	set of origin-destination (OD) pairs
q_k^i	travel demand of origin-destination $k \in Q$ at phase i
W_k^i	set of paths between origin-destination $k \in Q$ at phase i
h_w^i	flow on the path $w \in W_k^i$ (at phase i)
$\delta_{a,w}^i$	Link-path incidence index, equals 1 if link $a \in A \cup A'$ belongs to path w at phase i and 0 otherwise
M	a sufficiently large value
Z	value of the objective function: the total delay over all the construction phases

$$d_{p4}^i - y_{p2}^i \leq 0 \quad i = 1..n \quad (5)$$

$$\sum_{j=1}^{n-1} d_{px}^j - M \cdot y_{px}^n \leq 0 \quad px \in A' \quad (6)$$

$$y_{px}^i - y_{px}^{i+1} \leq 0 \quad px \in A', i = 1..n-1 \quad (7)$$

$$d_{px}^i, y_{px}^i \in \{0,1\} \quad px \in A', i = 1..n \quad (8)$$

$$\left\{ \begin{array}{l} \min_{x^i} \sum_{a \in A \cup A'} \int_0^{x_a^i} t_a(x) dx \\ s.t. \\ \sum_{w \in W_k^i} h_w^i = q_k^i, \quad k \in Q, \\ x_a^i = \sum_{w \in W_k^i} h_w^i \cdot \delta_{a,w}^i, \quad a \in A \cup A', \\ x_a^i \leq M \cdot y_{ax}^i, \quad a \in A', \end{array} \right\} i = 1..n \quad (9)$$

Equation (1) describes the upper-level goal of minimizing the total delay over n construction phases. Constraints (2)...(7) establish the Gantt chart that is the phased-structure of the construction and the prerequisite requirements as well as the concurrent constructions. There exists a subtle synergy between constraints (2) and (3) which work hand-in-hand to institute the phased-structure. If the project px is still not completed up until phase i (i.e. $\sum_{j=1}^{i-1} d_{px}^j < D_{px}$), then, according to constraint (2), the respective project cannot be considered completed (i.e. $y_{px}^i = 0$). In the meantime, constraint (3) is apathetic to the value of y_{px}^i (i.e. $y_{px}^i \in \{0,1\}$) thanks to the presence of ε a tiny positive value. All in all, when $\sum_{j=1}^{i-1} d_{px}^j < D_{px}$ holds true, constraints (2) and (3) result in $y_{px}^i = 0$. Similarly, when all the construction phases (or phases in short) of a project are completed until phase i (i.e.

$\sum_{j=1}^{i-1} d_{px}^j = D_{px}$), although constraint (2) becomes indifferent (i.e. $y_{px}^i = \{0,1\}$), this time, constraint (3) ascertains that the respective project at the respective phase is already completed and operational (i.e. $y_{px}^i = 1$). The auxiliary variables are “defined” and devised to link up D_{px}, d_{px}^j in constraints (2) to (8). When the optimization model (1)-(9) is solved, in one go, all the variables simultaneously get their own values such that the organic linkage between the trio ($d_{px}^i, y_{px}^i, D_{px}$) is upheld. Constraint (4) enforces that the consumption of the resources by the projects at the respective phases stays below the allocated resource (note that constraint (4) can be replicated to include a variety of different resources such as skilled/unskilled workforces, machinery equipment, budgets etc). Constraint (5) represents one of the prerequisite cases in which construction of the project p4 will never begin (at phase i) unless project p2 is already completed and operational. Similar prerequisite constraints can be added to the formulation if need be.

Constraint (6) ensures that no construction is started unless it is certain that construction of the respective project will be completed by the end of the construction period.

Constraints (7) and (8) ensure that both decision variables are binary and there is consistency among y_{px}^i . Formulation (9) represents a series of user-equilibrium TAP corresponding to n phases within which the last constraint ensures that projects corresponding to no-build decisions ($y_a^i = 0$) are excluded from the traffic assignment.

The main point of this article is constraints (2-8) in which a complicated Gantt chart has been formulated as a series of binary and linear constraints and the methodology is tailored to real-life networks.

As a concrete depiction of the formulation, consider Fig 1 showing a construction plan consisting of 6 projects, divided into 8 phases as well as some prerequisite requirements (colored red). It also shows availability of the resource and consumptions. As can be seen (in the Gantt chart), due to the limited resource, some projects (p1, p3, and p5) have seen pauses in their constructions, and project p6 was not selected at all. In the two last rows of the Gantt chart, the corresponding decision variables of project p3 (as an example) are shown, in which this project is under construction only during phases 1 and 4. Hence from phase 5 onward, project p3 is assumed completed and operational. Each column or phase of the Gantt chart represents a time duration or a construction time, that could be days, weeks or even months. Each phase can have its own time duration mandated by the projects’ nature. However, to have a consistent analysis, the unit of time duration must be normalized to the “least common multiple” of all the projects time duration. For example, if there exist four projects with construction times of say 4, 2, 10, 6 months, the unit of time duration will be 2 months. Therefore, each column of the Gantt chart represents time duration of 2 months as a construction phase. By doing this, the Gantt chart will be normalized and for each phase, a traffic assignment is carried out. Though in most regional transportation model, there exist multiple time periods with different OD pattern, for the sake of simplicity we run a typical peak hour assignment. However, in the proposed methodology, across different phases, the OD pattern can vary as shown in the case of Chicago. Nevertheless, expanding traffic assignment to multiple time periods is a trivial task for a modeler to be included in this methodology.

As defined in the formulation and shown in Fig 1, the resource consumptions by the projects are assumed to be identical across the phases. Nevertheless, it is easy to rectify such a shortcoming by expanding the workflow. To this end, consider for instance project p3 with two phases, associated with 4 and 5 units of the resource. Let us replace project p3 with two successive projects (say p3-1 and p3-2) each having their own resource’s requirements (4 and 5 units of resources respectively) in such a way the project p3-1 yields zero benefits whereas the project p3-2 inherent the entire benefit of

the project p3 upon completion (in other words, it is easy to split a project in more interdependent projects, using the prerequisite constraints, as a convenient way to distinguish between resource needs for various phases).

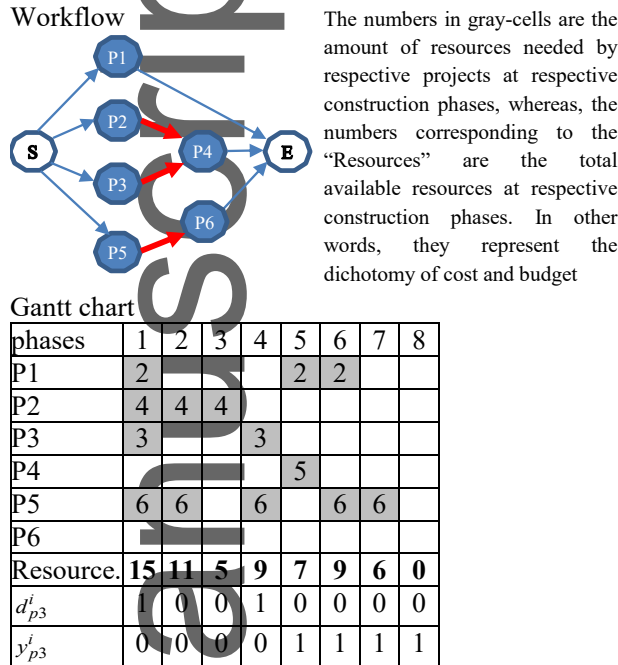


Fig 1 An illustration of a workflow and a Gantt chart for the road construction projects

4. A SOLUTION ALGORITHM

The fact that the RSPSP is a bilevel problem is enough to make it a NP-hard problem (Ben-Ayed and Blair, 1990). The consensus is to decompose a bilevel problem into two sub-problems corresponding to two upper and lower bound values of the objective function. Such schemes quickly become paralyzed when they are confronted with a large-sized network. The authors have developed a hybrid method for general bilevel problems consisting of machine-learning (a supervised learning technique) and optimization (an integer programming problem) tailoring to real-life applications called ML-OP (Bagloee et al., 2018) which is also employed here as a solution algorithm.

The main tenant of the ML-OP is to exploit the advantages of the two methods to tailor a practical solution NP-hard bilevel problems. There is a plethora of papers using the ML in engineering applications (Castillo et al., 2017; Oh et al., 2017; Rafiei and Adeli, 2015; Rafiei et al., 2016).

The algorithm initiates with a feasible integer solution specifying what to do in each phase (perhaps the do-nothing scenario, we call it a naïve solution or scenario) for which a series of TAPs corresponding to the phases are solved to calculate the objective function. In other words, given a feasible solution for the binary decision variables (d_{px}^i, y_{px}^i), the above problem (1..9) is solved and the value of the objective function Z is computed. These data (Z vs. d_{px}^i, y_{px}^i) are then used to train a multivariate linear regression model as a function of the decision variables:

$$\bar{Z} = \sum_{px \in A'} \sum_{i=1}^n a_{px}^i \cdot d_{px}^i + \sum_{px \in A'} \sum_{i=1}^n b_{px}^i \cdot y_{px}^i \quad (10)$$

where $\bar{Z}$ is a linear approximation of the original objective function (i.e. Z , total delay, Eq. (1)) and a_{px}^i, b_{px}^i are parameters to be calibrated. In other words, given a set of training data (Z vs. d_{px}^i, y_{px}^i) an estimation of the parameters (a_{px}^i, b_{px}^i) is sought. The calibration process is equivalent to solve for a quadratic minimization of the gap between Z and $\bar{Z}$ subject to several linear constraints which is mathematically a convex, tractable and benign optimization problem. Given the widespread applications of the regression models, there exist some highly efficient and purposely customized solution methods available in off-the-shelf soft wares. In this study, we make use of “regress” a user-friendly module available in MATLAB.

The idea of replacing the cumbersome objective function of design problems such as the NDP with a more tractable function known as a meta-model or surrogate function is not new. Wismans et al. (2014) address the NDP using dynamic traffic management measures formulated as a multi-objective NDP. They use the response surface methods (RSM) as an approximation method in combination with the genetic algorithm. Moreover, the same concept (surrogate model) has also been used for simulation-based optimization aiming to combine information from a traffic simulation model and a parsimonious trust-region optimization framework. Osorio and Bierlaire (2013) apply this method to solve a fixed-time traffic signal control problem for a subnetwork of the Lausanne city center, which has shown promising results.

The next step is to arrive at a new feasible and earnestly better binary solution such that this new solution should minimize the objective function. Hence we set up an integer linear programming (ILP) problem as follows. The regression model which is a linear function of decision variables is assumed to be a valid approximation for the original objective subject to all the binary constraints of the original problem (and most importantly, excluding the sub-problems (9)) as follows:

$$\min_{d_{px}^i, y_{px}^i} \bar{Z} = \sum_{px \in A'} \sum_{i=1}^n a_{px}^{i,k} \cdot d_{px}^i + \sum_{px \in A'} \sum_{i=1}^n b_{px}^{i,k} \cdot y_{px}^i \quad (10)$$

$$\begin{aligned} \sum_{(px,i) \in Y1^j} y_{px}^i - \sum_{(px,i) \in Y0^j} y_{px}^i &\leq |Y1^j| - 1, \\ Y1^j &= \{(px,i) \mid y_{px}^{i,j} = 1\}; \\ Y0^j &= \{(px,i) \mid y_{px}^{i,j} = 0\}, j = 1..k \end{aligned} \quad (11)$$

$$\begin{aligned} \sum_{(px,i) \in D1^j} d_{px}^i - \sum_{(px,i) \in D0^j} d_{px}^i &\leq |D1^j| - 1, \\ D1^j &= \{(px,i) \mid d_{px}^{i,j} = 1\}; \\ D0^j &= \{(px,i) \mid d_{px}^{i,j} = 0\}, j = 1..k \end{aligned} \quad (12)$$

(2) ... (8)

where constraints (11) and (12) ensure that a new set of binary variables (d_{px}^i, y_{px}^i) is always obtained. Note that $Y1^j, Y0^j$ represent variables (y_{px}^i) that take values of either 1 or 0, respectively and j denotes collections of all the respective solutions from the first iteration up to the previous iteration

k . Similarly, the same holds true for the $D1^j, D0^j$ and d_{px}^i . In fact, one of the main advantages of formulating such a complex problem using only binary variables (as decision variables) is to be able to use the above linear constraints (originally developed by Balas and Jeroslow (1972)) to avoid replicated solutions.

Furthermore, the above formulation consists only of several binary variables which are relatively easier to be solved compared to a general ILP or a mixed integer linear programming problem. The new binary solution (d_{px}^i, y_{px}^i) is then fed into the original problem (1)...(9) to calculate the corresponding total delay as the objective function Z . Any new solution (d_{px}^i, y_{px}^i and Z) contains valuable information about the original problem which is then used to re-estimate the parameters of the regression model.

The original problem RSRSP which is a mixed integer nonlinear problem (MINLP) is solved iteratively. First, the variables of the problem are divided into two sets, binary and continuous. Second, a mixed integer programming problem (MIP), as a sub-problem already derived from the RSRSP is solved to find a solution for the binary variables. Hence this binary solution is fed into the original RSRSP to transform it to a nonlinear programming problem (NLP, or simply a series of TAPs) which is easier to be solved compared to the original MINLP. In other words, the original MINLP (i.e. RSRSP) is divided into two sub-problems MIP and NLP to be solved alternatively in an iterative fashion. Note that, due to high similarity of the original MINLP and the sub-problem NLP (NLP is, in fact, MINLP when the binary variables are replaced with their corresponding solutions), when we are solving the NLP, we say, given a binary solution the MINLP is solved. This process carries on for a pre-specified number of iterations. At the end, the best solution is reported.

Obviously, the first regression starts with only one training record for which the corresponding calibration problem (i.e. minimization of the error between observed and estimated data) will have many optimal solutions for the regression parameters. However, we choose one calibration set out of many and proceed to the next iteration, where, the next calibration will be carried out with two training records and so on (note that the purpose of the early iteration is solely training, hence the algorithm is agnostics to the solutions, if presented by the regression). In fact, MATLAB, the software used for the regression, renders only one solution and a warning message that the number of training records is less than the number of parameters (as the number of iterations and hence the number of training records exceed the number of parameters, no more warnings will be reported).

In numerical evaluations, we show that several good solutions are found in early iterations. This complies with similar observations reported in the literature for the exact method. Bagloee et al. (2017a) have recently developed an exact method based on Benders decomposition for the NDP in which the optimal solutions are found in early iterations, whereas, the rest of the iterations strive toward closing a gap between the upper and lower bounds as a termination criterion. When dealing with a large-sized NP-hard problem, such a criterion (i.e. closing the gap between lower and upper bounds) is untenable. Thus, a user-specified maximum number of iterations (denoted by $k_{\max}$) is used as a termination criterion which depends on the preference of the user, the available computational power as well as the importance of the problem (however, in the numerical test we show that monitoring fluctuation of the solutions gives a convincing clue of what would be an appropriate estimation for the $k_{\max}$).

Though the solution algorithm bears some resemblance to the linear approximation methods, the linear approximation is a term used usually to refer to the methods in which all (or significant) elements of a problem are transformed into linear functions without omitting any components.

However, in the ML-OP a linear regression function replaces not only the objective function but also all the constraints related to the TAPs (1) and (9) which are dropped in the ILP.

As a recap, the algorithm of the proposed methodology is provided here.

Step 0: Set $k := 1$ as an iteration counter, $k_{\max}$ as a maximum number of iterations and $d_{px}^{i,k} = y_{px}^{i,k} = 0$ as an initial feasible solution for the binary variables.

Step 1: Given the current feasible solution $d_{px}^{i,k}, y_{px}^{i,k}$, solve the RSPSP (1..9) and return value of the objective function Z^k ,

Step 2: Given all the $d_{px}^{i,j}, y_{px}^{i,j}$ as well as their corresponding Z^j s compiled so far ($j = 1..k$), calibrate a new multivariate regression function $\bar{Z}^k$ of equation (10), which return $a_{px}^{i,k}, b_{px}^{i,k}$,

Step 3: Given the current binary solutions $d_{px}^{i,k}, y_{px}^{i,k}$ and calibrated parameters $a_{px}^{i,k}, b_{px}^{i,k}$, update the objective function (10) and constraints (11, 12) in the ILP problem,

Step 4: If $k < k_{\max}$ then set $k = k + 1$, solve the updated ILP problem and return a new feasible solution $d_{px}^{i,k}, y_{px}^{i,k}$, then go to Step 1. Otherwise, return the best solution found ($d_{px}^{i,m}, y_{px}^{i,m}, Z^m = \min\{Z^j\}, j = 1..k_{\max}$) and stop. ■

In fact given a feasible solution for the binary decision variables at iteration k (i.e. $d_{px}^{i,k}, y_{px}^{i,k}$), in Step 1, the original RSPSP (1..9) becomes merely a series of the TAPs. This process could be called a machine-learning method in the sense that invaluable results being collected from a series of all-out TAPs are used to train a black box (say the machine learning part) just to replace the original problem with a more intractable problem (10) as a guide to give us a new solution for the decision variables. However, this new solution is then fed into the original problem in which again a series of all-out TAPs are solved without any simplification and approximation.

5. REAL-LIFE CASE-STUDIES

We use large-sized data of the city of Winnipeg, Canada, and Chicago, U.S. for numerical tests (Bagloe 2017). The Winnipeg road network is comprised of 154 zones, 943 nodes and 3075 directional links and an hourly travel demand of 56,219 trips. Although in our formulation, it is a trivial task to call on a different travel demand for each construction phase, for the ease of illustration and without loss of generality, the same original travel demand matrix is assigned across the construction phases.

The Chicago network has 1790 zones, 11232 nodes, and 39098 directional links as well as an hourly travel demand for the base year of 1360434 trips. To show the generality of the methodology the travel demand is assumed to vary with .5% across successive construction phases.

We use a desktop computer with Intel_ Xeon _ 3.70 GHz and 64.0 GB RAM. The algorithm is coded with Visual Basic linked to MS-Excel and MS-Access as an interface and to save/retrieve database. The computer code is also synchronized with EMME 4 to solve the TAPs. The code also calls on MATLAB to solve the ILP and the regression using “intlinprog” and “regress” functions, respectively.

To solve a TAP a relative gap of .01 suffices practical purposes (INRO, 2017). However, to stay on the conservative side we use the rate of .001 for the Chicago and a very fine rate of .0001 (Boyce et

al., 2004) for the Winnipeg. For the Winnipeg, we use the Frank-Wolfe algorithm available in most of the commercial software including the EMME4 to solve the TAPs which takes approximately 3 seconds. Whereas, for the Chicago, given its unprecedented size we employ parallel-computation technique and a state-of-the-art and highly efficient method which will be discussed later. In both case-studies, we consider a pool of 20 new road projects to be selected and scheduled. The authors have proposed the concept of switch-link (Bagloee et al., 2017a), in which a project, as inextricable as a spaghetti interchange, in its entirety, is linked to a single link to be switched on, if the respective project is decided to be undertaken and completed.

5.1. Winnipeg

Figure 2 and Error! Reference source not found. present 20 projects as well as their respective number of construction phases where the entire constructions are planned to finish within 8 phases (e.g. for the first project, cost of 5 implies committing 5 out of 8 construction periods). Moreover, these projects are “wisely” selected to mainly complement several ring roads in the periphery of the central business district (CBD).

Table 2
The candidate projects

id	i-node*	j-node	traffic volume	cost
1	887	889	1761.04***	5
2	595	602	1600.24	3
3	513	595	1600.24	2
4	424	437	1233.9	1
5	420	592	774.524	6
6	301	1035	924.32	3
7	325	330	1555.34	2
8	739	774	683.708	2
9	297	1057	998.34	1
10	304	423	1420.27	1
11	551	610	1598.9	3
12	168	784	833.558	1
13	173	829	787.744	2
14	177	853	928.873	6
15	335	449	386.35	2
16	288	294	1349.26	3
17	327	424	562.273	1
18	299	1058	424.579	2
19	330	428	183.186	3
20	441	494	170.567	1

* Each row represents a two-way road with speed of 50 km/h and a capacity of 1700 vehicles per hour. The prerequisite condition for the Winnipeg case-study is as follows, construction of project 4 cannot start unless constructions of projects 2 and 3 are completed.

** Cost indicates the number of construction phases a respective project needs 1 unit of the resource or budget, for example, project 2 with the cost of 3 means, the project must be under construction in 3 construction phases. If the total budget available for a construction phase is 6, it means at this period, maximum 6 projects can be under construction. (note that for the sake of brevity with no loss of generality we assumed that at each construction phase, the projects that are under-construction, each consumes only 1 unit of the budget). Although the total cost is 50 units, a higher budget of say 70 units cannot necessarily guarantee selection of all projects for two reasons (i) the prerequisite conditions (ii) inefficient distribution of the budget over construction phases

***The traffic column (pertaining to the scenario where all the projects are completed) is there just to give a feeling of the level of traffic congestions and circulation.

There exists only one type of resources (called budget) available to the total amount of 10 units per phase excluding the final phase (this is to make sure that the last phase is the delivery phase in which no more construction is underway). In other words, total available budget throughout the construction periods is $70 = (8-1) \times 10$. Note that as indicated in Table 2, for the ease of representation, without loss of generality, the cost of construction per phase is identical to 1 unit (of the budget). For instance, the total number of phases for the first two projects are 5 and 3 which can also be interpreted as the total construction costs. As can be seen, given the number of projects (20) and the phases (8), we will have a total of $320 = 2 \times 20 \times 8$ binary decision variables, which can greatly represent large sized real-life examples.

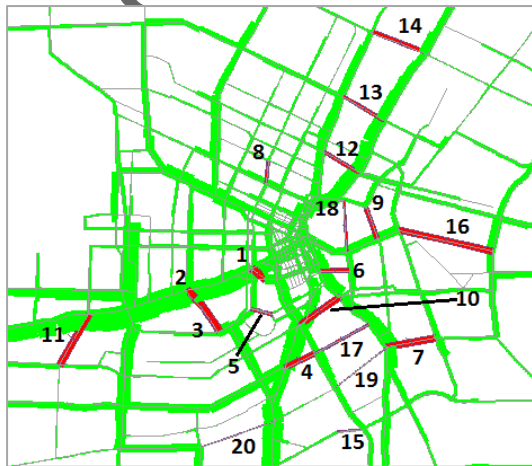


Fig 2 Winnipeg case study, 20 candidate projects

We run the algorithm for different budget levels. We first run it for 10% budget to get some insight about the degree of goodness of the solution. Note that 10% budget (i.e. budget of $1 = 10\% \times 10$) allows for only one project being under construction per phase for which we have already carried out an exhaustive enumeration to elicit the global optimum solution. Note that each run of the enumeration consists of solving a traffic assignment problem for 8 times (corresponds to 8 phases). We then run for 9 more budget levels of 20% to 100% with a pace of 10%, for 500 iterations and the results are presented in Table 3.

As can be seen, the algorithm managed to reach the global optimal solution for the first budget level (10%) during 5 hours computational time. The algorithm is also applied to the remainder of the budget levels as reported. In these cases, there is no clue of the optimal solutions, and hence it is not clear how good the final solutions are. To find some palpable sense of the quality of the final

solutions, we answer the following query: in the absence of any practical methodology how good is the proposed methodology?

Since the global solutions for such a large sized example is not known (except for the budget level of 10%), let us first randomly generate several feasible, greedy and informed solutions as “intuitive” solutions. We then compare these intuitive solutions with what our algorithm generates. The main idea is that, a good solution has many binary values 1, (note that, as mentioned before, the projects are defined wisely, to complement ring roads around the CBD, as no one wants to waste his/her investments when he/she puts forward a project as a proposal in the first place). We first populate the binary variables all with 1, then we start randomly switching them to 0 until we get a feasible solution. For each budget level, 10 solutions are generated and the averages are reported in Table 3 as “Intuitive”. As can be seen, the improvements that these greedy-intuitive solutions made to the doing nothing scenario vary from 0.4% to 1.1%, whereas, our algorithm managed to make these improvements more than double.

Although, these improvements seem minuscule, in fact, they are not. The benefit margins of transport projects are too slim while costs are onerous. For instance, a net benefit of millions of dollars investment into 10 grade-separation interchanges of a real-life example was found to be only 2% (in terms of traffic improvement), (see Table 6 of (Poorzahedy and Rouhani, 2007)). Therefore, even a substantial investment in the road expansion project could ultimately trickle down to a minuscule improvement of the traffic circulation depending on the size of the road network and other factors.

Table 3 also reports on the results of training a multivariate linear regression (a supervised learning algorithm): root-mean-square error (RMSE) and R-square. As can be seen, the multivariate linear regression provides a very tight approximation for the original objective functions. The high percentage of the R-square might be merely due to the fact that, a clear majority of the iterations (as shown in Figure 3) happens to be very close to each other when the algorithm converges. To this end, as an exercise, for the case of 50% budget level, we dropped all the records after iteration 176, where the U-shape ends, and calculated the R-square again. By doing so, the R-square was found 99.75% which is close to the overall R-square. In other words, the regression function obtained at the end, is also closely fit the U-shape records in early iterations where the variation of the objective value is high. This is an inexorable endorsement of enriching an optimization method with a machine learning technique for such a complex and difficult problem.

The nonlinearity feature of the RSRSP problem ultimately boils down to the traffic volumes of the individual roads. Variations of the traffic volumes of a clear majority of the roads with respect to different construction scenarios are negligible. This is because, much of the roads are local or arterial roads connecting trip generation and attraction points to the network which are captive traffic volume with not many alternative routes. In other words, when major changes happen to the road network, these changes do not necessarily ripple through the entire network. Therefore, the regression model as a parsimonious model can fit and capture these changes.

Computational times are also reported in the last two columns of Table 2. In more than half of the cases, the best solutions are found in the first half of the computations. One may find 10 hours CPU time (for the budget level of 90%) unaffordable. Nevertheless, with respect to the highly strategic nature of the problem when the fate of a significant investment (say millions of dollars or so) is at stake, 10 hours or even days is quite acceptable. In other words, computational time must be treated as a relative concern depending on the importance of a problem.

Figure 3 exhibits variations of the values of the objective functions over successive iterations. First, the objective value improves slightly and abruptly, but then it deteriorates and at a certain point promptly decreases again and then stays at a certain level with fewer variations. To explain this

behavior, it is important to note that the problem itself is a minimization problem whereas the objective value Z is a large positive value. The algorithm starts with a zero-out objective function. In early iterations, those variables who are assigned value 1 (when the ILP is solved in Step 4), will then be given some large positive coefficients in the regression process (Step 2). Hence in the next iteration, when the minimization problem is solved (the ILP in Step 4), the variables with large positive coefficients in the objective function will not be selected and instead the variables with the coefficient of zero will be selected (and will be given the value of 1). Therefore, the algorithm is forced to first exhaust all the variables with zero coefficients in $\bar{Z}$ which results in a significant deterioration of the objective value until a certain iteration. This process depicts a U-shape. This is a healthy setup to get the algorithm to explore the entire solution space. This U-shape part can be regarded as a training stage. When the training is consolidated, the algorithm generates some good performing solutions around optimal solutions (though we cannot certainly say that they are global optimal solutions). This is a very significant observation that can be utilized to specify a convincing clue about the maximum number of iterations. As the result indicates, a modeler must first let the algorithm finishes the training stage and carry on for some additional iterations (say twice the length of the training stage), before deciding to terminate the iterative process.

The solution method used in this study has a strong similarity with surrogate or meta-models when the original objective function is being constantly approximated based on a training dataset being accumulated over successive iterations. One may ask whether the approximation could be being skewed to a locally optimal solution. As a short answer, there is no proof to say otherwise. However, (i) the authors have numerically shown that the ML-OP method can find globally optimal solutions for a similar problem (the discrete NDP) very efficiently (Bagloee et al., 2018). (ii) The efficacy of the ML-OP has also been investigated when it was compared to an exact algorithm, that is Benders decomposition and branch and bound. Moreover, as shown in Figure 4, the incremental growth of the computational time over successive iterations is not exponential, it is, in fact, a linear trend (iii) it is also important to note that we are dealing with an NP-hard problem to which the ML-OP is a heuristic (not analytically exact) method. Hence what is expected from a heuristic method is to render some good (locally optimal) solutions in a timely manner, (v) in light of the existing literature, to the best of our knowledge, there exists no other solution for the optimization of the Gantt chart.

Similarly, we compared the results to an intuitive solution for two reasons (i) given the size and complexity of the problem (NP-hard) it is impossible to identify the best possible solutions (except for one case, the budget level of 10%, for which we carried out an exhaustive enumeration). Hence, in the absence of this information and in the absence of any better solution one conceivable way is to compare the results with an “intuitive” solution. (ii) Given the fact that we are dealing with an NP-hard, we need to adjust our expectation, that is, arriving at the best solution in real-world examples is mathematically proven to be unattainable. Having said that the aim is to arrive at good (locally optimal) solutions as shown in Figure 3 when the objective values converge.

Table 3
Numerical results of all the budget levels

Budget (%)	Binary string of the solution at the last phase	Best solution found objective function	Improvement (%)**		RMSE	R-square (%)	CPU% to reach best solution	CPU time (hrs)
			Intuitive	Best				
10	00000000110110100000	6588844.7*	0.403	1.048	656.1	99.04	84	4.06

20	0110001011011010000 0	6566697.5	0.565	1.381	435.2	99.86	27	5.49
30	0111000011011011110 1	6545218.3	0.701	1.704	493.01	99.94	26	7.14
40	0111001111111011110 1	6532498.3	0.781	1.895	1075	99.91	81	9.33
50	0111001111111011111 1	6525275.4	0.831	2.003	674.8	99.07	39	9.74
60	0111011111111111111 1	6520191.6	0.866	2.079	1040.3	99.85	40	9.82
70	0111011111111111111 1	6514462.6	0.901	2.165	956.0	99.54	60	9.85
80	0111011111111111111 1	6511399.2	0.925	2.211	916.1	99.96	83	10.01
90	0111111111111111111 1	6508407.8	1.010	2.256	811.6	99.88	89	10.1
100	0111111111111111111 1	6506331.6** *	1.100	2.288	798.5	99.91	51	9.7

* it is the global optimal solution based on an exhaustive enumeration,

** “Intuitive” is a greedy solution, and the “Best” is the best solution found in 500 iterations

*** the total delay of the do-nothing scenario is 6658652.1.

Not only the sheer size of the computational time is of the highest importance, its growth over successive iterations is also paramount. This would help us to understand if an algorithm is applicable to mega large examples. The computational time of all the budget levels are depicted in Figure 4. As is evidenced, there exists an almost linear trend which is a promising sign, in the sense the algorithm is amenable to gigantic problems.

The slight upward bend in the CPU time is correlated to the bottom of the U-shape segment, that is up until that point, solving the ILP (at Step 4) is relatively an easy job. However, when the algorithm proceeds to the second half of the U-shape (as it is forced to exhaust the variables with zero coefficient), the algorithm needs to weight up more binding and competing constraints which adds slightly to the CPU time.

The sensitivity of the methodology to the initial solution is also an important issue to be investigated. In a complicated situation with a multitude of projects and inextricable prerequisite, it might be too cumbersome to draw up a feasible scenario just by looking at the projects and weighing up resources, budgets and projects' requirement (in some cases there could be no such a feasible scenario to start with). Hence we did kick-start the methodology merely by a do-nothing scenario and let the algorithm to search for feasible scenarios in the next iteration. In case there exists no such a feasible scenario, the algorithm will be able to report it which is an invaluable information. However, one may ask what if the algorithm starts with an intuitive solution other than the do-nothing scenario, would the outcome be highly sensitive to the starting solution? To answer this question, we run the algorithm for the Winnipeg network and budget of 50% in which the starting feasible solution is the last solution found in the previous run when we run the algorithm with the do-nothing scenario (this is a very good solution closed to the local optimal solution). Figure 5 shows the variation of the results in which little disparity is seen. In other words, the proposed methodology is not sensitive and dependent on the starting point which is a good news and speaks of the reliability of the methodology.

Finally, Figure 6 depicts the workflow and corresponding completion status of the best solution found for the budget level of 80% (i.e. the Gantt chart). Considering Figure 6, there are four observations worth noting: (i) As can be seen, projects are constructed in the earliest possible periods such that no more construction is committed in the 7th period, even though enough budget is provided (ii) The only prerequisite condition was upheld, that is, construction of project 4 started only when constructions of projects 2 and 3 are completed. (iii) The tablet of the workflow clearly indicates which projects deserve to be constructed first (prioritization) and concurrent constructions of the projects in different phases are allowed.

As noted before, the crux of the problem is the scheduling part (the Gantt chart) based on which one can infer the prioritization order of the projects. (iv) Furthermore, as shown in Figure 6, despite the abundance of the budget, not all the projects have been selected for construction. By a visual inspection of Figure 6, it is easy to show that by shifting some projects from early phases further to the end, the two left-out projects (1 and 5) can also be accommodated into the Gantt chart. In other words, despite the abundant budget, the methodology managed to identify a better solution without these two projects (Bagloee et al., 2013). It is important to note that one of the advantages of the ML-OP is being parsimonious and frugal when it comes to the details of the original problem. In fact, the ML-OP can be treated as a black box with two inputs (i) the binary decision variables and (ii) their corresponding real values of the objective function which are the results of solving a series of traffic assignment problems (TAPs). Hence one can replace the TAP part with any other traffic simulation model such as dynamic assignment model, traffic assignment with elastic travel demand, stochastic assignment etc. Accordingly, in the next section, we provide a new case-study based on the large-sized network of Chicago in which the travel demand varies over the successive construction phases.

5.2. Chicago

The case-study of Chicago is added here to discuss the challenges and issues that are faced when addressing real-life problems in the industry. Given our engagement with the industry we must highlight some facts, (i) unfortunately some commercial software being used by practitioners and metropolitan areas are not computationally efficient. For example, using newly developed algorithms, the traffic assignment of a city of the size of Chicago with a fine relative gap (0.01) would take around a couple of minutes (see Table 2 in (Xie and Xie, 2016)), whereas it may take hours when using some of these software. To this end, we are using EMME 4, one of the most efficient software (see algorithm SOLA in (Florian and Morosan, 2014) available in EMME 4) based on which solving a typical TAP for the network of Chicago takes only 70 seconds (using 8 threads parallel computation). (ii) the model available in some metropolitan are more than a traffic assignment (there could be some elements of demand forecasting to prepare travel demand matrices before running a traffic assignment). That is why these models may take hours to run. Like the Winnipeg, we consider 20 new road projects wisely selected to ease traffic of the CBD, for which Table 4 provides a list of the corresponding switch-links. For the ease of referencing and sake of brevity, we carried over the same phase/budget/cost setup from the Winnipeg to the Chicago's case-study. The travel demand varies with the construction phase that is to take the long duration of the construction phases into account (note that one may set a construction phase of one year, it is too long and travel demand will not be the same for the next construction phase). For the do-nothing scenario, total delay over 8 phases is found to be 362295924 minutes. If it is divided by the total travel demand 11075849 trips, it would be

equivalent to an average of 32 minutes per trip. We run the algorithm for the budget level of 50% which takes 63 hours to terminate at iteration 500. The best solution is found at iteration 455 of 347422960 mins which is equivalent to 4 % improvement. The variation of the objective values, as well as the CPU time over the iterations, are illustrated in Figure 7. As can be seen, the variations of the objective value consisting of the U-shape part followed by congestion is similar to that of the Winnipeg. The same applied to the CPU time which was found to be a linear trend (since solving the TAP is much costlier compared to the Winnipeg case, we do not see the Winnipeg's slight upward bend).

Finally, to further enhance the realism of the methodology, in addition to the expense/budgetary constraint (4), we include a constraint (14) representing limited skilled labor force and machinery which vary over different phases (time steps) denoted by L^i and l_{px} is the amount of labor/machinery that project px needs.

$$\sum_{px \in A} l_{px} \cdot d_{px}^i \leq L^i \quad i = 1..n$$

(14)

To challenge the model we introduce a tight labor of 6, 4, 3, 5, 4, 3 and 5 unit for construction phases of 1 to 7 respectively. We assume the projects' labor requirements are identical to the costs presented in Table 4. With all things being equal, the numerical results indicate that the objective function significantly degraded (347422960 vs. 348212856). Details of all the inputs/outputs and EMME's codes are provided in the GitHub (Bagloee 2017).

6. CONCLUSION

We addressed selection, prioritization and scheduling constructions of the road projects using a heuristic method. We formulated the Gantt chart with some linear binary constraints while considering fidelity of the model by including three major features (i) phased structure of the roads' construction, (ii) prerequisite conditions, (iii) concurrent constructions of different projects. Moreover, the issue of the interdependency of the projects' benefits has been thoroughly addressed via proposing a bilevel problem, in which a series of traffic assignment problems (TAPs) are solved. Our main motive for this study is rooted in our engagement with the industry, especially in developing countries with massive investments into infrastructures for which there exists no systematic decision support system. To tailor to the industry-size case-studies we employed a heuristic method recently developed by the authors for the general bilevel problems which is a hybrid machine-learning technique and optimization (ML-OP).

For numerical tests, we used two large-sized case-studies, Winnipeg and Chicago. The methodology demonstrates a vivid potential to converge to locally optimal solutions. With respect to the CPU time, solving TAPs especially for networks of the size of Chicago is highly costly. To this end, we used the Second-Order Linear Approximation (SOLA) a new and highly efficient algorithm and a parallel computation technique. Nevertheless, the CPU time was found to be of the order of a couple of hours which is not unaffordable for the following reasons. For one, the unprecedented size of the case-

studies which are intentionally chosen to challenge the applicability of the methodology to real-life cases. Second, one needs to take the importance and strategic level of the problem itself into account. For example, when a problem is associated with millions or billions of dollars at stake, it will be affordable to wait for days or months to get a better solution. Third, unlike the traffic assignment which is a routine task to be solved frequently for traffic analysis studies, the scheduling problem (RSRSP) is, in fact, a one-off problem to be solved once, even if, it comes at the cost of say 24 hours computation. Forth, nowadays, for some traffic consultants, a weekend computational time has become an acceptable norm, to which solving the RSRSP can be easily accommodated.

Limitations of this methodology are: (i) we assumed a fixed, statistic and deterministic traffic model, (ii) there was only one mode of transport, public transport was missing, (iii) the merit of projects were assessed based on the travel time, a single objective function, (iv) short-term variation of drivers' behaviour due to construction operations, such as temporary road closure and detour plan was not considered.

This work can be further enhanced in two main avenues. (i) consideration of project related to the public transport such as locations of the stations, new metro lines is a missing dot for which the sub-problem (9) must be replaced with a multimodal and multiclass traffic model (Bagloee et al., 2012). To this end, although there exist some usually computationally expensive methods, the authors are working on a combined model based on a complementarity method (Bagloee et al., 2017b). (ii) in light of ever surging concerns of the environment and social aspects of transportation, moving toward a sustainable transportation is becoming a necessity such that traffic improvement should not be regarded as a solo-criterion for the projects' appraisal. Therefore expanding the proposed formulation to a multi-objective optimization is a worthwhile effort.

Nevertheless, it is important to note that the RSRSP, as it stands, is too complicated for which there exists no solution. This paper seeks credit for being first in two areas: (i) The first to formulate the Gantt chart with all its details and entirety for road projects using some "linear binary" constraints. Being linear and binary would enable us to embed the Gantt chart into the framework of a variety of optimization problem as a module. For example, the authors are working on a linkage project (a joint project between university and industry) in which recovery operations of a disaster-hit area is scheduled using the concept of the Gantt chart. (ii) The methodology is tailored to real-life problems.

Acknowledgment

The authors are grateful to Prof. Adeli and seven anonymous reviewers for their meticulous, insightful, and constructive comments. The first author acknowledges the constructive debates that he had with Tarek Abu Gharbieh and Nadeem Siddiqui from Road and Transport Authority (RTA), Dubai Government, United Arab Emirates, as part of the R956 project which became one of the motives of this study.

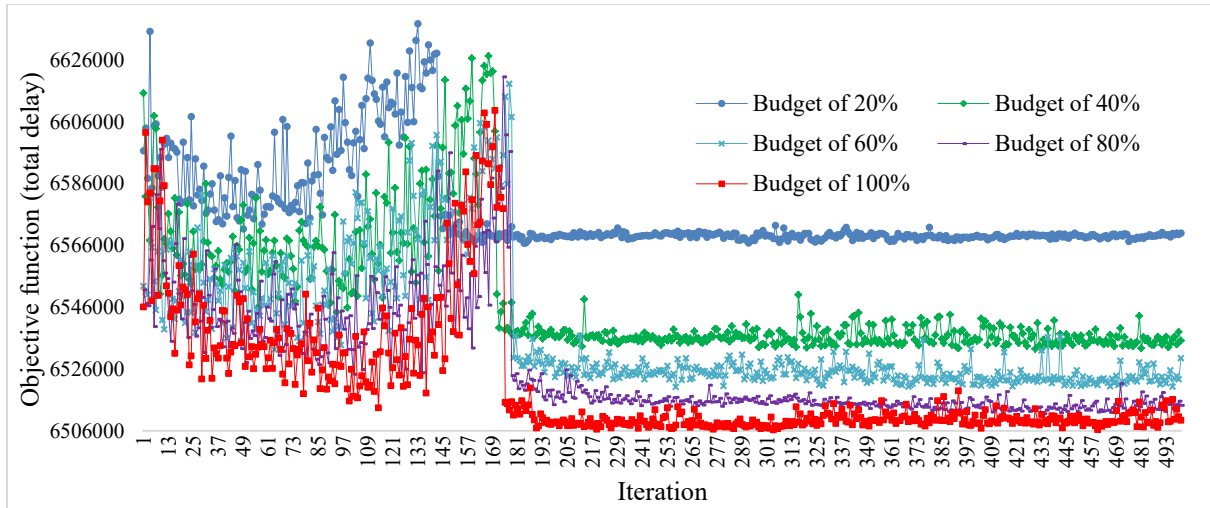


Fig 3 variations of the objective functions over successive iterations

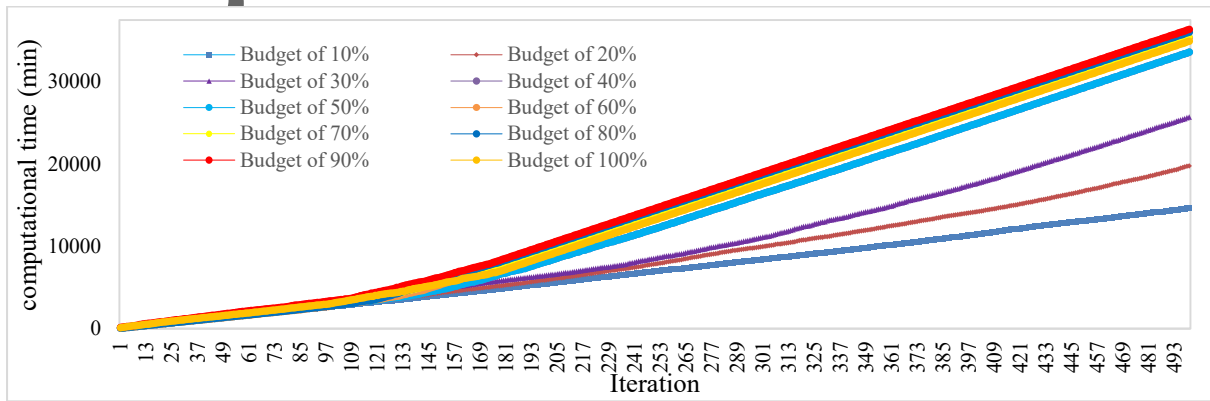


Fig 4 Computational time over successive iterations pertaining to the Budget level of 80%

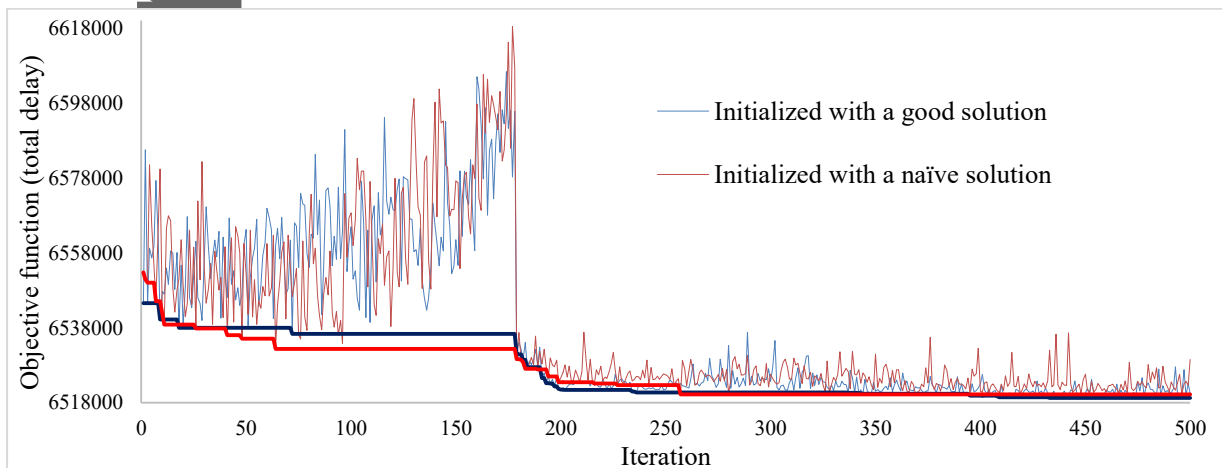


Fig 5 Winnipeg case-study, two runs for the budget level of 50% with different initial scenario

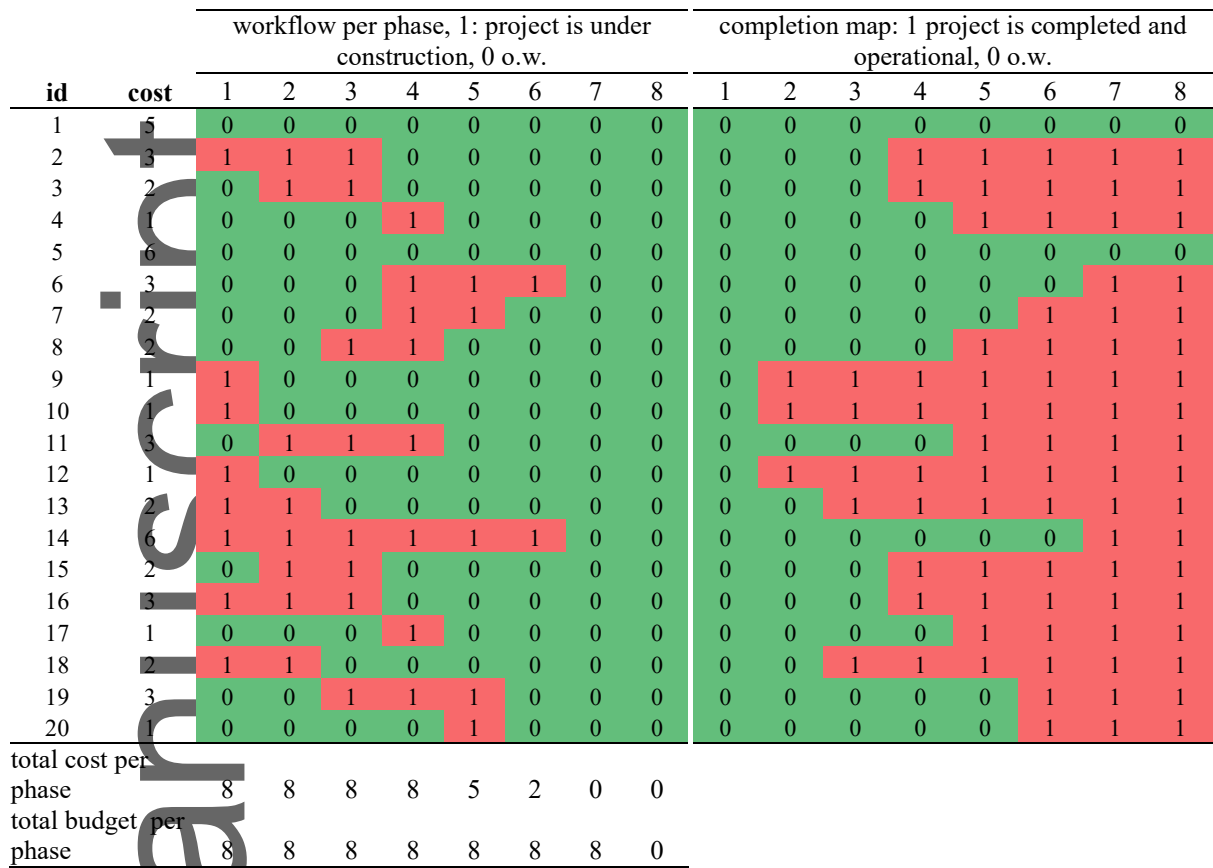


Fig 6 Gantt chart: work flow and completion status of the best solution found for budget level of 80%

Table 4
Chicago's case-study, list of switch links and corresponding costs

i	12984	12985	12987	12989	12991	12993	12995	12997	12999	13001	13003	13005	13007	13009	13011	13013	13015	13017	13019	13021
j	12983	12986	12988	12990	12992	12994	12996	12998	13000	13002	13004	13006	13008	13010	13012	13014	13016	13018	13020	13022
cost	5	3	2	1	6	3	2	2	1	1	3	1	2	6	2	3	1	2	3	1

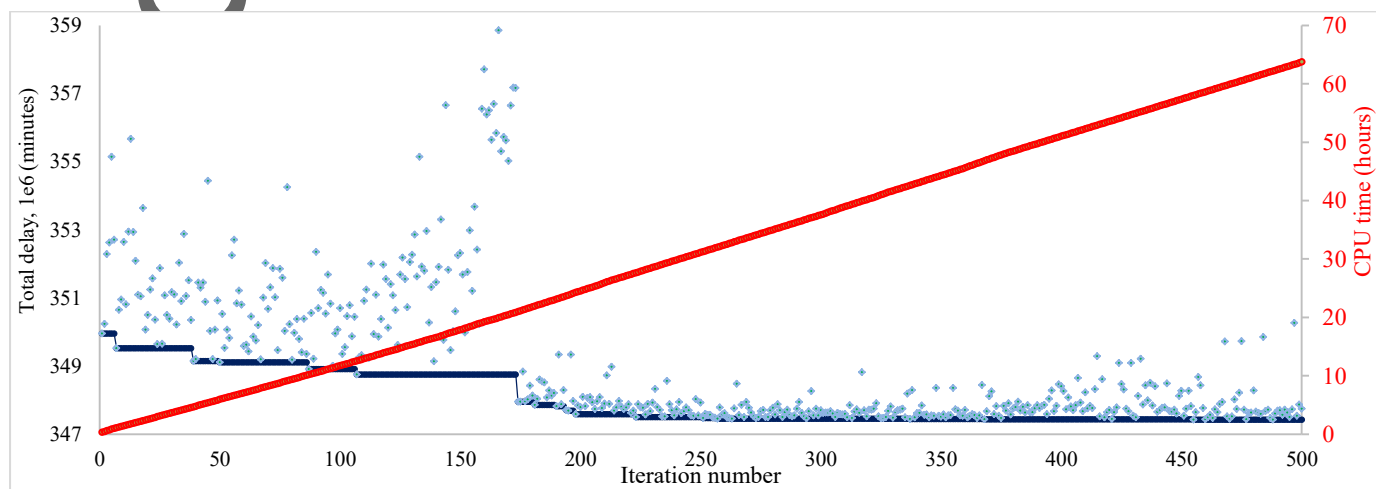


Fig 7. Chicago case-study, variations of the objective value and CPU time over successive iterations

REFERENCES

- Adeli, H., Karim, A., 2003. *Construction scheduling, cost optimization and management*. CRC Press.
- Bagloee, S.A., 2017. Optimization of the Gantt chart, road construction projects, Datasets of Winnipeg and Chicago, https://github.com/saeedasadibagloee/Gantt_chart, Melbourne, Australia.
- Bagloee, S.A., Asadi, M., 2015. Prioritizing road extension projects with interdependent benefits under time constraint. *Transportation Research Part A: Policy and Practice* 75, 196-216.
- Bagloee, S.A., Asadi, M., Richardson, L., 2012. Methodology for Parking Modeling and Pricing in Traffic Impact Studies. *Transportation Research Record: Journal of the Transportation Research Board* 2319, 1-12.
- Bagloee, S.A., Asadi, M., Sarvi, M., Patriksson, M., 2018. A hybrid machine-learning and optimization method to solve bi-level problems. *Expert Systems with Applications* <https://doi.org/10.1016/j.eswa.2017.11.039>.
- Bagloee, S.A., Ceder, A., Tavana, M., Bozic, C., 2013. A heuristic methodology to tackle the Braess Paradox detecting problem tailored for real road networks. *Transportmetrica A: Transport Science* 10, 437-456.
- Bagloee, S.A., Sarvi, M., Patriksson, M., 2017a. A Hybrid Branch-and-Bound and Benders Decomposition Algorithm for the Network Design Problem. *Computer-Aided Civil and Infrastructure Engineering* 32, 319-343.
- Bagloee, S.A., Sarvi, M., Patriksson, M., Rajabifard, A., 2017b. A mixed user-equilibrium and system-optimal traffic flow for connected vehicles stated as a complementarity problem. *Computer-Aided Civil and Infrastructure Engineering* 32, 562-580.
- Bagloee, S.A., Tavana, M., 2012. An efficient hybrid heuristic method for prioritising large transportation projects with interdependent activities. *International Journal of Logistics Systems and Management* 11, 114-142.
- Balas, E., Jeroslow, R., 1972. Canonical cuts on the unit hypercube. *SIAM Journal on Applied Mathematics* 23, 61-69.
- Ben-Ayed, O., Blair, C.E., 1990. Computational Difficulties of Bilevel Linear Programming. *Operations Research* 38, 556-560.
- Boyce, D., Ralevic-Dekic, B., Bar-Gera, H., 2004. Convergence of traffic assignments: how much is enough? *Journal of Transportation Engineering* 130, 49-55.
- Castillo, E., Grande, Z., Mora, E., Xu, X., Lo, H.K., 2017. Proactive, Backward Analysis and Learning in Road Probabilistic Bayesian Network Models. *Computer-Aided Civil and Infrastructure Engineering* 32, 820-835.
- Cochran, M.A., Pyle, E., GREENE, L.C., Clymer, H., Bender, A., 1971. Investment Model for R and D Project Evaluation and Selection. *IEEE Transactions on Engineering Management*, 89-&.
- Florian, M., Morosan, C.D., 2014. On uniqueness and proportionality in multi-class equilibrium assignment. *Transportation Research Part B: Methodological* 70, 173-185.
- Fox, G.E., Baker, N.R., Bryant, J.L., 1984. Economic models for R and D project selection in the presence of project interactions. *Management science* 30, 890-902.

- Gear, T.E., Cowie, G.C., 1980. A note on modeling project interdependence in research and development. *Decision Sciences* 11, 738-748.
- INRO, 2017. EMME 4 v 4.3.3, *EMME4 User's Guide*, 4.3.3 ed, Montreal, Quebec, Canada.
- Johnson, E.L., Kostreva, M.M., Suhl, U.H., 1985. Solving 0-1 integer programming problems arising from large scale planning models. *Operations Research* 33, 803-819.
- Jong, J.-C., Schonfeld, P., 2001. Genetic algorithm for selecting and scheduling interdependent projects. *Journal of waterway, port, coastal, and ocean engineering* 127, 45-52.
- Karim, A., Adeli, H., 1999a. CONSCOM: An OO construction scheduling and change management system. *Journal of Construction Engineering and Management* 125, 368-376.
- Karim, A., Adeli, H., 1999b. OO information model for construction project management. *Journal of Construction Engineering and Management* 125, 361-367.
- Kim, B.J., Kim, W., Song, B.H., 2008. Sequencing and scheduling highway network expansion using a discrete network design model. *The Annals of Regional Science* 42, 621-642.
- Lim, T.K., Yi, C.Y., Lee, D.E., Arditi, D., 2014. Concurrent construction scheduling simulation algorithm. *Computer-Aided Civil and Infrastructure Engineering* 29, 449-463.
- Lo, H.K., Szeto, W.Y., 2009. Time-dependent transport network design under cost-recovery. *Transportation Research Part B: Methodological* 43, 142-158.
- Nemhauser, G.L., Ullmann, Z., 1969. Discrete dynamic programming and capital allocation. *Management Science* 15, 494-505.
- Oh, B.K., Kim, K.J., Kim, Y., Park, H.S., Adeli, H., 2017. Evolutionary Learning based Sustainable Strain Sensing Model for Structural Health Monitoring of High-rise Buildings. *Applied Soft Computing*.
- Osorio, C., Bierlaire, M., 2013. A simulation-based optimization framework for urban transportation problems. *Operations Research* 61, 1333-1345.
- Ponz-Tienda, J.L., Pellicer, E., Benlloch-Marco, J., Andrés-Romano, C., 2015. The fuzzy project scheduling problem with minimal generalized precedence relations. *Computer-Aided Civil and Infrastructure Engineering* 30, 872-891.
- Poorzahedy, H., Rouhani, O.M., 2007. Hybrid meta-heuristic algorithms for solving network design problem. *European Journal of Operational Research* 182, 578-596.
- Rafiei, M.H., Adeli, H., 2015. A novel machine learning model for estimation of sale prices of real estate units. *Journal of Construction Engineering and Management* 142, 04015066.
- Rafiei, M.H., Khushefati, W.H., Demirboga, R., Adeli, H., 2016. Neural Network, Machine Learning, and Evolutionary Approaches for Concrete Material Characterization. *ACI Materials Journal* 113.
- Sarma, K.C., Adeli, H., 2001. Bilevel parallel genetic algorithms for optimization of large steel structures. *Computer-Aided Civil and Infrastructure Engineering* 16, 295-304.
- Senouci, A.B., Adeli, H., 2001. Resource scheduling using neural dynamics model of Adeli and Park. *Journal of Construction Engineering and Management* 127, 28-34.
- Sinha, A., Malo, P., Deb, K., 2017. A Review on Bilevel Optimization: From Classical to Evolutionary Approaches and Applications. *IEEE Transactions on Evolutionary Computation*.
- Szeto, W.Y., Jaber, X., O'Mahony, M., 2010. Time-Dependent Discrete Network Design Frameworks Considering Land Use. *Computer-Aided Civil and Infrastructure Engineering* 25, 411-426.
- Tao, X., Schonfeld, P., 2005. Lagrangian relaxation heuristic for selecting interdependent transportation projects under cost uncertainty. *Transportation Research Record: Journal of the Transportation Research Board* 1931, 74-80.

- Tao, X., Schonfeld, P., 2006. Selection and scheduling of interdependent transportation projects with island models. *Transportation Research Record: Journal of the Transportation Research Board* 1981, 133-141.
- Tao, X., Schonfeld, P., 2007. Island Models for Stochastic Problem of Transportation Project Selection and Scheduling. *Transportation Research Record: Journal of the Transportation Research Board* 2039, 16-23.
- Taplin, J., Qui, M., 2001. Using GA to Optimize the Selection and Scheduling of Road Projects. *The Practical Handbook of Genetic Algorithms Applications*, 99-134.
- Wang, S., Schonfeld, P., 2012. Simulation-Based Scheduling of Mutually Exclusive Projects with Precedence and Regional Budget Constraints. *Transportation Research Record: Journal of the Transportation Research Board* 2273, 1-9.
- Weingartner, H.M., 1966. Capital budgeting of interrelated projects: survey and synthesis. *Management Science* 12, 485-516.
- Wismans, L., Van Berkum, E., Bliemer, M., 2014. Acceleration of solving the dynamic multi-objective network design problem using response surface methods. *Journal of intelligent transportation systems* 18, 17-29.
- Xie, J., Xie, C., 2016. New insights and improvements of using paired alternative segments for traffic assignment. *Transportation Research Part B: Methodological* 93, 406-424.
- Zhang, H., Tam, C., Li, H., 2006. Multimode project scheduling based on particle swarm optimization. *Computer-Aided Civil and Infrastructure Engineering* 21, 93-103.