SHORTEST PATHS IN FREIGHT MULTIMODAL NETWORKS WITH NON-ADDITIVE IMPEDANCES: A PRACTICAL APPROACH

Vittorio Marzano, Corresponding Author
Assistant professor
Dipartimento di Ingegneria Civile, Edile e Ambientale
Università di Napoli Federico II
Via Claudio 21, 80125, Napoli (Italy)
Tel: 0039-081-7683935; Email: vmarzano@unina.it

Andrea Papola
Associate professor
Dipartimento di Ingegneria Civile, Edile e Ambientale
Università di Napoli Federico II
Via Claudio 21, 80125, Napoli (Italy)
Email: papola@unina.it

Fulvio Simonelli
Assistant professor
Dipartimento di Ingegneria
Università del Sannio
Palazzo Bosco Lucarelli, Corso Garibaldi, 82100 Benevento (Italy)
Email: fulvio.simonelli@unisannio.it

Roberta Vitillo
PostDoc Researcher
Dipartimento di studi aziendali e quantitativi
Università degli Studi di Napoli Parthenope
Via Generale Parisi 13, 80132, Napoli (Italy)
Email: roberta.vitillo@uniparthenope.it

Daniela Tocchi
PhD student
Dipartimento di Ingegneria Civile, Edile e Ambientale
Università di Napoli Federico II
Via Claudio 21, 80125, Napoli (Italy)
Email: daniela.tocchi@unina.it

Word count: 6,026 words text + 5 tables/figures x 250 words (each) = 7,276 words
Submission Date: July 29, 2016.
Revised Version Submission Date: November 15, 2016.
ABSTRACT

The paper aims at tackling the issue of calculating shortest paths in multimodal freight networks, a key challenge for the presence of non-additive impedances (e.g. travel times, costs, fares) and also for the inherent heterogeneity of freight transport supply. Conceptually, since non-additive impedances can be handled only with explicit path enumeration, the proposed approach looks for a more tractable and less computationally demanding explicit path enumeration, based on two main steps. The former is the creation, starting from the monomodal networks for each freight mode, of so-called macrolinks representing direct connections between origins and destinations of that mode: this allows turning non-additive impedances in the initial monomodal network into additive impedances in the macrolink-based monomodal network, obviously at the price of increasing the number of links in the network. The latter is that all macrolink-based monomodal networks can be coupled together in a multimodal macrolink-based network, leveraging and slightly modifying earlier contributions in the literature, to achieve an effective model for the entire multimodal freight network. The proposed approach is applied to a very large real-size network representing the entire Euro-Mediterranean freight transport supply, demonstrating its capability to account for the effects of truck drivers’ stop/rest time regulations (typically non-additive) on port choice and maritime service choice. Results show how the proposed approach actually improves the capability of calculating realistic shortest paths, with important policy implications, preserving a tractable computational burden.

Keywords: shortest path, non-additive freight impedances, multimodal freight supply.
1. BACKGROUND AND MOTIVATION

Calculating shortest paths in multimodal freight networks is important for a wide range of applications, including policy-making and governance – see for instance (1) and (2) – and also for supply chain and transport operations’ optimization, e.g. (3). In general, both synchronic and dyachronic approaches can be applied, as argued by (4) amongst others. However, focusing attention on decision support systems at national and international level for transport policy and governance, multimodal freight supply models largely resort on synchronic networks (5), and also embed remarkable simplifications to preserve full integration with equilibrium-based demand-supply interactions and to reduce calculation times. In fact, albeit acceptable for some types of policies, such simplifications might lead to considerable modelling errors in key policy applications, for instance in the identification of the catchment areas of ports and in the analysis of competition amongst freight modes.

In this respect, a first necessary characteristic of a multimodal freight supply model is its capability to model impedances associated to terminal nodes (i.e. nodes allowing transfer between modes), which can change depending on the connected modes and on the type of multimodal leg (e.g. first access, transshipment): for instance, a container transshipment operation in a port is normally associated with times and costs different from a container import/export operation. A second necessary characteristic is the capability to account for the remarkable heterogeneity of freight transport options (own account/third party, type of commodity, type of vehicle, type of loading unit…), which can lead to substantially different performances. A noteworthy approach capable to handle both issues is the NODUS model by (6) and (7). NODUS exhibits two key features: first, a specific topological representation of terminals – based on “exploding” all within-terminal connections between modes – enables differentiation of impedances for all possible transfers between modes; second, “virtual links” are created to model connections between terminals, being each virtual link representative of a specific freight service, e.g. characterized by different types of vehicles and/or different costs/prices. NODUS has been applied in some contexts, mainly for the analysis of freight elasticities and for the location of new freight terminals.

Mainly, two key phenomena should be taken into account when modelling freight supply, that is the presence of multiple impedances to minimize, and their non-additivity. Indeed, from a fairly general viewpoint, four types of impedances can be taken into account: travel times, costs (i.e. in the light of the operators of the freight transport service), fares (i.e. in the light of the users of the freight transport service), and generalized costs. Generalized costs are normally given by a linear combination of costs and other factors (e.g. travel times, reliability) opportunely harmonized through monetary coefficients, e.g. value of travel time saving, value of reliability, see for instance (8). In addition, a key issue in modelling freight supply is dealing with the presence of non-additive impedances. By definition, a non-additive impedance cannot be associated with any specific links in the networks, thus preventing calculation of path impedances as sum of impedances of all links belonging to that path. Two main examples of non-additive freight impedances are non-linear freight fares, whose unit value (e.g. €/km or US$/mile) usually decreases by distance, and the regulation on driving times and rest periods for truck drivers, leading to non-additive total travel times, see for instance (9). Travel costs can be non-linear as well with respect to freight flows, as a result of the superposition of two contrasting effects, the economies of scale on one hand and the congestion on the other hand.
Some authors have proposed shortest path algorithms trying to accommodate the aforementioned issues. The minimization of a linear combination of two impedances (the so-called bi-criterion shortest path) is proposed by (10), whilst several researches considered the effect of nonlinear disutility in the optimization problem. In (11), for instance, an algorithm to solve a bicriterion shortest path problem with continuous monotone functions is provided. In (12) it is suggested a heuristic method for solving the non-additive shortest path problem with general continuous cost function. In (13), the general non-linear cost function is first approximated with a piecewise linear counterpart and then each linear subproblem is sequentially solved. Another recent solution consists of a parametric search method which works on the bicriterion shortest path problem, solving a series of subproblems, i.e., the parameterized subproblems (14).

A particular sub-class of non-additive impedances is called sub-additive: by definition, an impedance \( \bar{i} \) is sub-additive if it can be decomposed into an additive component \( i_{\text{add}} \) and a non-additive component \( i_{nadd} \), such that \( i_{nadd} \) is a non-decreasing function of \( i_{\text{add}} \). This means that, given any two paths indexed by 1 and 2, the following holds:

\[
i^1_{\text{add}} < i^2_{\text{add}} \Rightarrow i_{nadd}(i^1_{\text{add}}) < i_{nadd}(i^2_{\text{add}}) \Rightarrow i^1 = i^1_{\text{add}} + i_{nadd} < i^2 = i^2_{\text{add}} + i_{nadd} \tag{1}
\]

The key advantage of the sub-additivity property is that, thanks to (1), the shortest path with respect to the impedance \( \bar{i} \) is still the shortest path with respect to the sole additive component \( i_{\text{add}} \). In other words, it can be determined by calculating the shortest additive path with respect to the additive component \( i_{\text{add}} \) via a traditional shortest path algorithm (e.g. Dijkstra), and then adding in post process the corresponding non-additive component \( i_{nadd} \) in order to calculate its overall impedance.

A very relevant example of sub-additive impedances in modelling freight supply is represented by travel times: indeed, all regulations on truck drivers’ rest and stop times are such that, given the net road driving time \( t_{\text{road}} \), the mandatory additional rest/stop time \( t_{\text{road}} - t_{\text{road}} \) is a non-decreasing function of the net road driving time \( t_{\text{road}} \), i.e. equation (1) holds. In this respect, the work by (16) and the subsequent application by (17) leverage exactly this property to derive a modification of the Dijkstra’s algorithm capable to account for rest/stop times in a monomodal road freight network. Similarly, also the structure of many freight service fares is such that the unit fare \( p_u(d) \) (e.g. US$/mile or €/km) is a decreasing function of the distance \( d \), in a way such that the overall cost is subadditive. That is, given two paths 1 and 2 with length \( d_1 \) and \( d_2 \), their respective fares \( p_1=\frac{d_1}{p_u(d_1)} \) and \( p_2=\frac{d_2}{p_u(d_2)} \) are such that \( d_1 < d_2 \Rightarrow p_1=\frac{d_1}{p_u(d_1)} < \frac{d_2}{p_u(d_2)} = p_2 \), i.e. the property (1) still holds. Thus, being the distance an additive impedance, the shortest fare path in a monomodal network can be still calculated simply by calculating the shortest distance path and then calculating the corresponding fare, which will be the minimum possible thanks to the subadditivity property.

The key issue for the purposes of the paper is that the sub-additivity does not hold anymore in multimodal networks, i.e. when a path may include a sequence of different modes, each with specific non-additive impedances. In other words, one may take advantage of the sub-additivity property only if all links and paths in the network are characterized by the same impedance \( \bar{i} \). By way of example, Figure 1 provides a simple counterexample of a road-maritime freight multimodal network with sub-additive impedances (driving stop times) only on the road network: it is easy to recognize that the shortest additive time path (bold red in the middle of Figure 1) is actually not the shortest overall path (bold red in the bottom of Figure 1).
Figure 1 – Shortest time path in multimodal networks: a counterexample (network structure: top; shortest time additive path: middle; shortest time overall path: bottom)
This example also clarifies that, if the objective of the modelling is for instance to understand the catchment areas of ports or the flows on maritime services, not accounting for the non-additivity of freight impedances might lead to significant biases in policy making (in the example, in the throughput of ports C and D and in the maritime flows between C and D).

In such cases, explicit path enumeration is required to calculate the shortest path with respect to a non-additive impedance. Since the number of paths is normally very large, even for medium-sized networks, proper heuristics should be applied. For instance, one might adopt a random generation algorithm, whose steps at the generic iteration are: (a) sampling additive link impedances from a normal distribution; (b) calculating the shortest additive path with respect to that impedance, and (c) calculate the total path impedance by adding the non-additive components. A sufficiently large variance of the sampling distribution will increase the chance of detecting good candidate paths, however at the price of increasing substantially the computational burden.

Alternatively, (18) reproduced approximately the impact of road rest times in multimodal freight networks using NODUS by clustering o-d pairs by distance band and applying multi-class assignment with separate cost functions by distance band. However, a proper clustering should be based on travel time bands rather than on distance bands, in accordance with the above; also, this approach does not allow capturing the complex effects on non-additivity along the entire transport chain.

Thus, to the authors’ knowledge, there is no availability of an effective approach capable to deal with calculation of shortest paths in multimodal freight networks with non-additive impedances. This paper aims at filling in this research gap, proposing a modeling approach for the calculation of shortest paths in freight multimodal networks consistent with all the freight requirements mentioned above and, at the same time, parsimonious in terms of calculation times, thus consistent with the requirements by the inclusion into an operational decision support system.

The paper is structured as follows: Section 2 illustrates the proposed approach; Section 3 provides a real-network example; finally, Section 4 draws conclusions and research prospects.

2. FORMULATION OF THE PROPOSED APPROACH

This section describes the proposed approach for the calculation of shortest paths in multimodal freight networks with respect to a non-additive impedance. Freight supply performances highly depend upon a wide range of factors, related e.g. to the nature of good, the type of vehicle, the business relationships, and so on. Thus, a proper segmentation of the supply model should be adopted, developing a specific multimodal freight model for each freight segment: in the following we will refer to a single segment, i.e. the notation will not include any specific segment-related superscripts/subscripts. However, in principle, the proposed approach can be coupled with NODUS, that is virtual parallel links for each freight segment between two terminals could be created as well.

Conceptually, since non-additive impedances can be treated with explicit path enumeration and calculation of shortest path amongst enumerated paths (as recalled in Section 1), the problem is how to make this explicit path enumeration more tractable and less computationally demanding. For this aim, the approach proposed in the paper is based on two main steps:
• starting from the monomodal networks for each concerned freight mode, proper *macrolinks* are created, each representing a direct connection between an origin-destination pair of that mode; this allows associating macrolinks with non-additive impedances related to that origin-destination pair. As a result, non-additive impedances in the initial monomodal network are turned into additive impedances in the macrolink-based monomodal network, obviously at the price of increasing the number of links in the network;

• then, all macrolink-based monomodal networks can be coupled together in a multimodal macrolink-based network by applying the NODUS framework by (6)-(7), with a slight modification on road connections in order to avoid a side-effect in accounting for the driver’s rest times, as it will be shown later.

The two aspects are discussed in the following sub-sections.

**2.1 Building a macrolink-based monomodal freight network**

Let be $M$ a set of available freight modes indexed by $m \in 1...n_m$, comprising the road mode and other non-road modes. Each mode $m$ is represented by a “real” monomodal graph $\Gamma_m \equiv \{N_m, L_m\}$ being $N_m$ and $L_m$ the sets of nodes and links respectively. Without loss of generality, the set of nodes associated with the road mode encompasses a subset $C$ of centroids representing traffic analysis zones (TAZs). Furthermore, within the study area, a set $T$ of specific loading/unloading points, called terminals, with cardinality $n_\tau$, allows transfers across modes. Each terminal $\tau \in T$ is associated with a set $M_\tau \subseteq M$ of $n_m \tau$ modes, and for each mode $m \in M_\tau$ a corresponding node $\tau_m \in N_m$ exists. The union of all terminal nodes $\tau_m$ for the mode $m$ leads to the set $T_m$ of all terminals of mode $m$, with cardinality $n_\tau m$.

In the following, without loss of generality, attention will be focused on a specific impedance. If the impedance is additive, then there is no need to build a macrolink-based monomodal network, and the “real” monomodal graph can be used directly in the multimodal network. If the impedance is sub-additive, in accordance with the above, the original “real” graph for the mode $m$ should be replaced with another graph $\Gamma^*_m \equiv \{N^*_m, L^*_m\}$ in which $N^*_m$ contains only origins and destinations and $L^*_m$ all related macrolinks, being a macrolink $l^* \in L^*_m$ a direct connection between $o$ and $d$ with $o,d \in N^*_m$.

Specifically, the set of nodes $N^*_m$ includes the set of all nodes belonging to relevant o-d pairs for the mode $m$. In general, for the road mode, o-d pairs will include both terminals and centroids as origins and destinations, i.e. $N^*_\text{road} \equiv C \cup T_{\text{road}}$, whilst for non-road modes normally only terminals will be origins and/or destinations, i.e. $N^*_m \equiv T_m \ \forall m \neq \text{road}$. However, this is not a strict requirement.

In terms of macrolinks $L^*_m$, a distinction between road and non-road modes should be introduced. For the road mode, the connection between o-d pairs depends only on the presence of road infrastructures. For the other modes, the connection between o-d pairs (normally terminals, in accordance with the above) depends not only on the presence of infrastructures, but also on the availability of concerned freight services. Thus, building the macrolinks means basically building all connections between terminals depending on the available services. In the case of direct services, a macrolink directly connecting the two concerned terminals suffices. In the case of
services calling at various terminals (e.g. a maritime container service calling at various ports), let
\( \tau_1 \ldots \tau_n \in T_s \) be the sequence of those terminals, where \( T_s \) is the set of terminals of service \( s \), each
terminal \( \tau \in T_s \) should be connected with every other terminal \( \tau' \in T_s \) with \( j \neq i \) so as to associate all
possible connections amongst terminals within \( T_s \) with the appropriate sub-additive impedances,
i.e. leading to \(|T_s| \cdot (|T_s| - 1)/2 \) two-way macrolinks.

By way of example, Figure 1 embeds two maritime services, i.e. a direct service between ports A
and C and a service calling at ports A-B-D-E in sequence: the former will be represented by a
single macrolink, the latter by 6 two-way macrolinks connecting respectively ports A-B, A-D,
A-E, B-D, B-E, D-E. Obviously, this implies that a pair of terminals will be connected potentially
by many macrolinks: in this respect, a simplification of the network can be achieved as follows. In
the process of building macrolinks for a given service, one might encounter a pair of terminals
already connected by a macrolink representing a previously processed service. In this case, one has
to check whether the additional service is ameliorating the impedance associated with the already
existing macrolink: if so, the macrolink is associated with the new impedance, without adding a
new macrolink. This opens also an interesting possibility: in general, there might be multiple
impedances associated with a link/macrolink (e.g. time, costs, fares, waiting times/frequency).
Thus, the analyst might be interested in calculating all other path impedances associated with the
shortest path with respect to a given impedance: for this aim, when updating the impedance of a
macrolink in the sense above defined, the analyst should also replace all other impedances. This
would also enable calculating as many shortest paths as the number of concerned impedances, and
then comparing all these paths based on the values of all related impedances. Notably, the side
effect of this approach is that, since the same physical leg of the service is represented by multiple
disjoint links in the macrolink network, a specific post-process is required if a capacity-dependent
assignment (either for accounting for congestion or for capacity constraints) is performed based on
this modified network. For instance, in the example of Figure 1, the physical movement of freight
between ports A and C is split amongst macrolinks A-C and A-D.

Overall, each macrolink connecting two nodes \( o \) and \( d \) with mode \( m \) – either road or non-road –
will be associated with an impedance calculated on the “real” network of mode \( m \). This means, in
the case of sub-additive impedance, following the approach recalled in Section 1.

It always occurs \(|N_m| > |N^*_m|\) and \(|L_m| < |L^*_m|\), that is turning non-additive impedances in the real
digraph \( \Gamma^*_m \) into additive impedances in \( \Gamma^*_m \) through macrolinks comes at the price of a substantial
increase in the number of links, equal at most to \( n^m_{od} (n^m_{od} - 1) \) being \( n^m_{od} \) the number of
origin/destination nodes for the mode \( m \). Thus, to limit the maximum number of macrolinks,
connections with terminals with a total travel time leading at least to a rest period (i.e. > 8h) can be
excluded. Another rule-of-thumb may be enabling connections between terminals only if their
distance exceeds a minimum threshold (e.g. rail services are enabled only for distances higher than
300 km).

### 2.2 Building a multimodal freight network based on macrolinks

Once built the macrolink-based graph \( \Gamma^*_m = \{N^*_m, L^*_m\} \) for each mode \( m \), a multimodal graph can
be built by connecting all macrolink-based monomodal graphs using a slight modification of the
NODUS approach, as illustrated in the following Figure 2.
In formal terms, each mode $m$ using the terminal $\tau$ is associated with two nodes, an “entering node” $\tau_{in,m}$ and an “exiting node” $\tau_{out,m}$. Therefore, each terminal $\tau \in T$ is modelled by means of a number of nodes equal to $2 \times n_{m,\tau}$, possibly connected by the following types of links for each generic mode $m$:

- the backward star of the entering node $\tau_{in,m}$ includes all macrolinks of mode $m$ going to the terminal $\tau$. Specifically, if the mode is road, those macrolinks are either of type $\tau_{in,road} - \tau_{out,m}$, i.e. they represent the overall shortest paths from each origin $o \in C$ to the terminal $\tau$, or of type $\tau'_{out,road} - \tau_{in,road}$, being $\tau'$ another terminal connected by road. These links are capable to represent so-called road “land bridges”, e.g. in Europe a multimodal link between the Balkans and the Iberian Peninsula with a first maritime link in the Adriatic Sea, a road land bridge in Italy between an Adriatic and a Thyrrenian port, and a second maritime link between an Italian Thyrrenian port and an Iberian port. If the mode is not road, those macrolinks represent all services of mode $m$ arriving at terminal $\tau$ from other terminals of the same mode, i.e. links of type $\tau'_{out,m} - \tau_{in,m}$ with $\tau' \not\in T_i$, in accordance with the definition of macrolink reported in Section 2.1. Notably, in the approach by (6)-(7) the backward star of the entering node $\tau_{in,m}$ includes simply links of the monomodal network of mode $m$, possibly duplicated into “virtual links” to account for different freight demand services;

- the forward star of the exiting node $\tau_{out,m}$ is made up by all the concerned links of mode $m$ leaving from terminal $\tau$. In accordance with the above, if the mode is road, those macrolinks are either of type $\tau_{out,road} - d$, i.e. they represent the overall shortest paths from the terminal $\tau$ to a given destination $d \in C$, or of type $\tau_{out,road} - \tau'_{in,road}$, with $\tau' \not\in T_{road}$, so as to represent road land bridge. If the mode is not road, those macrolinks represent all services of mode $m$ leaving from terminal $\tau$, i.e. of type $\tau_{out,m} - \tau'_{in,m}$ with $\tau' \not\in T_m$.

- the forward star of the entering node $\tau_{in,m}$ is defined as follows:
  - if the mode is road, it is made up by all links representing connections with non-road other modes, i.e. by all links of type $\tau_{in,road} - \tau_{out,m} \forall m \neq road \in T \tau$. Each of such links represents access from the road mode to other modes, and therefore all
related first access terminal impedances should be associated with those links.

Notably, the connection $\tau_{in,\text{road}} - \tau_{out,\text{road}}$ should be prohibited, so as to avoid undesired effects. The example of Figure 1 allows clarifying this point: following the proposed approach, the origin $o$ should be connected via specific road macrolinks to the destination $d$ and to each port, and all ports could be also possibly linked together with road macrolinks. Recalling Section 2.1, and restricting attention to the travel time as impedance, each macrolink is associated with the shortest total path, inclusive of non-additive impedances. For instance, drawing upon EU regulations, the macrolink $o$-$C$ will be associated with 8 hours of net driving time plus a 0.75 h of short stop time, the macrolink $C$-$D$ with 5 hours of net driving time and 0.75 h of short stop time, and the macrolink $o$-$D$ with 13 h of net driving time, 1.5 h of short stop time and 8 h of rest time. Thus, it is clear that if within terminal $C$ a direct connection between entering and exiting road macrolinks were created, a shortest path $o$-$C$-$D$ would be found in the multimodal macrolink-based network with an impedance of 13 h of net driving time and 1.5 h of short stop time, i.e. the rest time needed because of the 13 h of net driving time would be discarded, making the correct macrolink $o$-$D$ dominated.

- If the mode is not road, it is made up by three types of links, exactly reprising the rationale underlying NODUS. The first is a possible link representing the connection with the road mode, i.e. a link $\tau_{in,m}$-$\tau_{out,\text{road}}$ which represents egress from the mode $m$ towards the road mode, therefore it will be associated with all related impedances (e.g. final custom clearance times and costs in the case of a terminal container). The second is a within-mode $m$ transshipment link of type $\tau_{in,m}$-$\tau_{out,m}$ which is associated with all within-mode transshipment impedances (e.g. in the previous example of the container terminal, only transshipment times and costs will be taken into account, without customs impedances). The third is a set of links of type $\tau_{in,m}$-$\tau_{out,m'}$ with $m'\neq\text{road}, m\in T_\tau$ representing the between-modes transshipment not involving road (e.g. from sea to rail in a container terminal), again with specifically designed impedances.

- the backward star of the $\tau_{out,m}$ node includes all links from the entering nodes of all modes connected with $m$, either with a first access link, a within-mode transshipment link and/or a between-modes transshipment link. Obviously, if the mode is road, the backward star of the node $\tau_{out,\text{road}}$ cannot encompass the link $\tau_{in,\text{road}} - \tau_{out,\text{road}}$ for the same reason explained above.

Overall, the proposed approach is depicted at a glance in the flowchart in Figure 3. As a result, it is important to underline that the proposed modified network with macrolinks contains only additive costs, which however include all non additive impedances in each monomodal network, and therefore standard shortest path algorithms can be applied for calculation of shortest paths. In addition, in principle, this network can also be used for standard assignment procedures applied in transport engineering (1). In that respect, in accordance with the above, it should be only noted that once performed the assignment on the macrolink, the resulting flows should be reported back on the single monomodal links of each single monomodal network. Notice also that one might not consider in the macrolink-based multimodal network the direct connection between centroids via the road mode: this will force the shortest path algorithm to look necessarily for a multimodal route.
Thus, this allows achieving a main improvement over the existing approaches, including NODUS, that is the capability to account explicitly and not approximately for the effects of driving rest and stop times, which can have a considerable impact on modelling the catchment areas of ports and the flows on maritime services. The price for that, i.e. the additional computational complexity, will be discussed in the next section, together with other implementation considerations.

It is also important to notice that the proposed approach fails in only one situation, specifically when considering travel time as impedance it could happen that the shortest path identified through the approach described in Figure 3 is not actually the overall shortest time path. The example of Figure 1 can be used to provide a counterexample: the shortest path identified in the bottom of the Figure 1 would actually be not correct if the shortest road path from the origin to the embarking port C were 9 hours instead of 8 hours. In that case, the regulation would imply an 8 hours stop time, whilst the maritime leg C-D would allow only 7 hours of recovery time. However, it is easy to circumvent this issue: once applied the proposed approach, it suffices to add a post-process control to detect all such incorrect shortest paths, and replace them by new paths obtained by repeating the proposed approach on a modified macrolink network whose impedances are augmented to 8 hours minimum (i.e. to a full recovery time) all non-road macrolinks.

Finally, in terms of computational burden, building macrolinks for each mode requires running shortest paths on monomodal networks, which is not time-consuming, whilst the creation and the storage of the macrolink-based network may take time and be also memory consuming. In this respect, specific well-known data structures can be adopted in order to obtain an efficient data storage, and parallelization might be also used to speed up calculation times.
3. EXAMPLE ON A REAL NETWORK

This section demonstrates an application of the proposed approach to a real-size network, in order to quantify its computational burden in a challenging context and also show how it actually improves the capability of calculating realistic shortest paths. For this aim, the approach of Figure 3 has been coded in Microsoft Visual C# without parallelization, and tests have been run on an HP notebook Intel i7 equipped with 8 GB of RAM and a 7200rpm hard disk drive.

The real case study is at the Euro-Mediterranean scale, based on the supply model described in (19) and firstly implemented in (20) and (21). This is a challenging case study, with a study area encompassing 57 countries and 1508 NUTS3 zones, and four modes: road, sea, rail, inland waterways. Specifically, the “real” monomodal networks include 63109 links and 50870 nodes for the road mode, 28139 links and 2147 nodes for the maritime mode, 1041 links and 966 nodes for the inland waterways mode and 90259 links and 83462 nodes for the rail mode. In terms of relevant nodes, there are 1508 centroids representing a NUTS3 zoning of the study area, encompassing 57 Euro-Mediterranean Countries, 491 ports and more than 3000 rail terminals/ramps.

By way of example, results are reported with reference to the calculation of shortest paths using a combination of road and maritime modes. Specifically, the following has been considered:

- concerned freight segments refer to Ro-Ro transport, with combination of two different options: accompanied, unaccompanied (i.e. whether the driver is onboard the Ro-Ro vessel in the maritime leg of the journey, or not); 1 driver, 2 drivers. In particular, since by nature the feeder legs to/from ports and origins/destinations are normally carried out with just 1 driver for unaccompanied Ro-Ro transport, there are three feasible freight segments: (a) unaccompanied with 1 driver; (b) accompanied with 1 driver; (c) accompanied with 2 drivers.

- for each segment, three impedances are taken into account:
  - travel times, including drivers’ stops and rest time regulations, differing for the 1 driver/2 driver segments based on current EU regulations;
  - travel costs, i.e. the analysis is made in the light of the road freight service provider. Ro-Ro cost includes both vessel onboard costs (usually the sea fare is lower for unaccompanied trailers) and port terminal costs (usually higher for unaccompanied due to terminalisation costs). Notice also that, overall, travell costs can be decomposed in fixed costs, distance-dependent costs and time-dependent costs;
  - travel generalized cost, calculated based on a monetization of the travel time with a specific value of travel time saving (VTTS). Relying to the case of third-party transport, the VTTS monetizes the value of the truck, which is assumed 13 €/h for the tractor and 5 €/h for the trailer.

- two types of mode combinations are taken into account:
  - all-road transport: it does not require building macrolinks, because the sub-additivity property holds for a monomodal network (Section 1).
  - multimodal road-sea transport: it requires adopting the approach described in Section 2, leading to a multimodal macrolink-based freight network encompassing 1508 centroids and 491 ports, i.e. leading to more than 740000 road macrolinks and 115000 sea macrolinks.
For each of the three segments and for each of the two mode combinations, the following shortest paths have been calculated:

- shortest time path, then calculation of the corresponding costs and generalized costs;
- shortest cost path. Specifically, only the additive time-dependent cost components have been taken into account. Then, calculation of the corresponding times, the additional costs due to non-additive time components, and generalized costs;
- shortest generalized cost path. Only additive time-dependent cost components and additive cost components have been taken into account, and each link has been associated with a generalized cost. The shortest path has been then calculated, and the additional non-additive concerned costs and times calculated.

Overall, each o-d pair has been associated with the shortest generalized cost amongst the three previously described shortest paths: clearly, this is a heuristic sub-optimal approach. Calculation times are very acceptable for planning-oriented applications, with less than one hour for the worst case. By way of example, it is worth analysing some calculated shortest paths: specifically, Figure 4 refers to the Milan-Sevilla o-d pair, interesting because the road paths and the multimodal road-sea paths are almost parallel, thus with comparable distance, whilst Figure 5 refers to the Milan-Tirana o-d pair, wherein the road-sea freight multimodal paths have a remarkable advantage with respect to the all-road paths. For both examples, shortest generalized cost paths calculated with the approach described above are depicted.

Figure 4 – Example of freight shortest path in a multimodal network with non-additive impedances: Milan-Sevilla o-d pair
Firstly, the approach reproduces correctly port choice and maritime service choice, based on comparison with direct observations of “true” shortest paths on the above mentioned o-d pairs. In the example in Figure 4, it is worth noticing that there is no difference between 1 driver and 2 driver multimodal shortest paths. In addition, in both examples, since the road mode is faster than the maritime mode, the proposed approach reproduces correctly the tendency to shorten the maritime leg of the path for higher VTTS freight segments, i.e. accompanied Ro-Ro will tend to exhibit a shorter maritime leg with respect to the unaccompanied transport. Importantly, the capability to account exactly for the non-linear effect of truck drivers’ stop/rest regulations is the main feature of the proposed approach capable to model correctly its effects on port and maritime service choice. This is particularly evident in the example of Figure 5 wherein, since the maximum road leg without stop is roughly 9 hours for 1 driver and the double for 2 drivers, the accompanied Ro-Ro shortest route will call at the port of Ancona for 1 driver (Ancona-Milan takes less than 8 hours) and of Bari for 2 drivers (Bari-Milan takes in between 9 and 18 hours).

A second test, not reported here for the sake of brevity, has been conducted on the multimodal USA road-rail freight network, based on the Freight Analysis Framework data toolbox (22). A study area with 17 ports, 150 rail ramps and 7137 5-digit zip codes has been selected, spanning over the entire United States, and the shortest multimodal paths with rail-road mode combinations have been calculated, with a calculation times of 46 minutes and 33 seconds.
4. CONCLUSIONS AND RESEARCH DIRECTIONS

The paper has presented an operational approach for calculating shortest paths in multimodal freight networks, a key challenge mainly for the presence of non-additive impedances (e.g. travel times, costs, fares) and for the inherent heterogeneity of freight transport supply. The proposed approach is based on two main steps. The former is the creation, starting from the monomodal networks for each concerned freight mode, of proper macrolinks representing direct connections between origin-destination pairs of that mode: as a result, non-additive impedances in the initial monomodal network are turned into additive impedances in the macrolink-based monomodal network, obviously at the price of increasing the number of links in the network. The latter is that all macrolink-based monomodal networks can be coupled together in a multimodal macrolink-based network, leveraging and slightly modifying earlier contributions in the literature, so as to model effectively the entire multimodal freight network.

The proposed approach has been applied to a very large real-size network representing the entire Euro-Mediterranean freight transport supply. Results show how the proposed actually improves the capability of calculating realistic shortest paths, with important policy implications, however with tractable computational burden. By way of example, its capability to account for the effects of truck drivers’ stop/rest time regulations (typically non-additive) on port choice and maritime service choice has been also presented.

Many research prospects can be identified: first, the inclusion of the proposed approach into freight assignment procedures should be assessed and operationalized. Then, the extension to the case of fully non-additive cost functions embedding economies of scale and congestion effects should be tested on real networks. Finally, more efficient algorithms can be developed to reduce further the computational burden of the proposed approach.

ACKNOWLEDGMENTS

This research was carried out in the frame of Programme STAR, financially supported by UniNA and Compagnia di San Paolo, through the project SMURFS led by the first author.

REFERENCES


