Transportation Research Part E 81 (2015) 36–51

Contents lists available at ScienceDirect

Transportation Research Part E

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

Demand clustering in freight logistics networks

Rodrigo Mesa-Arango, Satish V. Ukkusuri *

Lyles School of Civil Engineering, Purdue University, 550 Stadium Mall Drive, West Lafayette, IN 47907, USA

ARTICLE INFO

Article history:
Received 22 December 2014
Received in revised form 5 May 2015
Accepted 4 June 2015
Available online 8 July 2015

Keywords:
Combinatorial auctions
Community detection
Freight and logistics
Network clustering
Stochastic effects

ABSTRACT

Demand clustering in freight logistics networks is an important strategic decision for carriers. It is used to incorporate new business to their networks, detecting potential economies, optimizing their operation, and developing revenue management strategies. A specific example of demand clustering is truckload combinatorial auctions where carriers bundle lanes of demand and price them taking advantage of economies of scope. This research presents a novel approach to cluster lanes of demand. Community detection is used to cluster the emergent network finding profitable collections of demand. Numerical results show the advantages of this method.

1. Introduction

Logistics clusters are increasing around the world following the example of successful models like Dubai, Panama, Rotterdam, Memphis, Sao Paulo, Singapore, among others (Yu et al., 2005; Boile et al., 2011; Sheffi, 2012, 2013). Complementarities and synergies contribute to the economic prosperity of the constituent firms (commodity production, storage, transportation, and other supporting activities). Logistics clusters are significantly important for companies that provide freight transportation services. Sheffi (2013) summarizes the main competitive advantages achieved by transporters providing services in these places. The high volume of freight between clusters generates larger shipments. Thus, economies of scale are achieved because the shipment unitary cost is lower for vehicles that are filled to capacity. Firms operating large vehicles, e.g., Valemax ships or double stacking trains, considerably benefit from these economies. When vehicles are not filled by single shipments, freight logistics companies can distribute costs by consolidating several shipments in facilities and vehicles, and, hence, achieving economies of density. Intermodal companies related to the maritime and railroad modes usually benefit from this activity. Additionally, for the trucking mode, less-than-truckload (LTL) companies take advantage of these features. The huge amount of freight entering and leaving logistics clusters reduces idling times and fosters economies of frequency. In general, these economies benefit companies in all modes. Last but not least, symmetric flows between clusters propitiate economies of scope by reducing the fraction of shipment unitary cost associated to empty repositioning. This considerably benefits all modes, especially those with fixed facilities within clusters, e.g., docks, terminals, stations, consolidation facilities, etc.

Governments recognize the economic importance of logistics clusters and increasingly provide incentives for firms to (re)locate into these facilities (Sheffi, 2013). However, this is a slow process. Sometimes it is not even an alternative for many shippers and carriers that face enormous relocation costs, off-shoring issues, and potential detriment of relationships with clients. Additionally, logistics clusters might not be a feasible option because they have not emerged naturally, they are not a

* Corresponding author. Tel.: +1 765 494 2296; fax: +1 765 494 0395.
E-mail addresses: rmesaa@gmail.com (R. Mesa-Arango), sukkusur@purdue.edu (S.V. Ukkusuri).

http://dx.doi.org/10.1016/j.tre.2015.06.002
1366-5545/

priority for local governments, or they are not suitable for unstable economic landscapes. In these cases, flexible firms with low use of specialized facilities have competitive advantages over other rigid modes. The challenge for these firms is developing operations that mimic the advantages of logistics clusters, increase revenues for transporters, and add value to their clients.

Truckload (TL) companies are the best example of flexible freight transportation carriers. Undeniably, TL is the most popular type of operation for the most popular freight mode: Trucking. This mode accounts for 29% of the for-hire-transportation market share. This value is higher than the joint share for the second and third modes, i.e., air (16%), and rail (8.0%) (USDOT, 2012). Setar (2013a, 2013b) estimates that TL accounts for 61% of the 2013 US general trucking industry revenue ($193.4 Billion). TL firms are considerably impacted by economies of scope (Caplice, 1996; Jara-Diaz, 1981, 1983; Mesa-Arango and Ukkusuri, 2013) and frequency (Sheffi, 2013) induced by empty trips resulting from freight imbalances.

Firms know that empty trips profoundly affect their economy. Companies like Best Buy, Coca-Cola Supply LLC, JB Hunt Transport, Johnson & Johnson, Walmart Stores, Inc, among others, have participated of the Empty Miles program (VICS, 2014) to share unused transportation capacity and reduce empty-trip inefficiencies (Belson, 2010). In 2009, the chain of department stores Macy's cooperated with shippers and carriers to reduce 1500 empty trips in the US. In average, they saved $25,000 transportation costs annually for each shared lane (VICS, 2009). JCPenney, another important department-store chain, shared 41,000 backhauls that saved them $8.1 Million between 2008 and 2009 (Andraski, 2010). Schneider National, the largest private TL carrier in North America, increased dedicated backhaul revenue by 25% on specific accounts thanks to this initiative (VICS, 2009). Unfortunately, empty trips are not rare for trucking operations. 25% of the 2010 truck-kilometers in Europe where traveled empty (De Angelis, 2011). Reduction of empty trips can significantly benefit society because they are related to serious externalities like emissions, traffic congestion, and wear of roads. The monetary savings obtained by Scheider National also saved them 5554 gallons of diesel fuel that eliminated 61.65 tons of carbon dioxide, 147.24 tons of articulate matter, and 1.47 tons of nitrous oxide. Similarly, JCPenny eliminated 9750 tons of CO_2 by utilizing 20% of its empty miles in 2009 (4 million miles) and 6% (1.3 million miles) in 2008.

Although empty trips can be reduced through collaboration, TL carriers can develop strategies to promote this behavior making them more attractive and profitable. The challenge is detecting and clustering synergetic lanes, i.e., lanes that minimize empty trips when operated together. Additionally from a revenue management perspective, the right combination of TL volume and price has to be considered in the development of profitable clusters of demand. However, prices and volumes add a new level of complexity to this problem. The high level of competition in the TL market makes the development of pricing strategies very difficult. Thus, carriers that look at market values when analyzing clusters realize that they vary significantly. Variations in the observed traffic volumes (Caplice and Sheffi, 2006) also occur in a symbiotic fashion. This happens for several reasons: seasonal changes (e.g., end of the year or harvests), forecasting errors, macroeconomic impacts (e.g., economic recessions or booms), network disruptions (e.g., inclement weather), among others. An approach that incorporates these sources of uncertainty can significantly benefit the development of demand clusters.

The concept of clustering has been approached in similar works. Bidding advisory models have been developed to bundle lanes in TL combinatorial auctions (CA) (Song and Regan, 2003, 2005; Wang and Xia, 2005; Lee et al., 2007; Chang, 2009; Huang and Xu, 2013; Xu and Huang, 2013, 2014; Kuyzu et al., 2015; Triki et al., 2014; Ergun et al., 2007). Additionally, geographic clustering has been used to reduce the computational complexity of vehicle routing problems (Bowerman et al., 1994; Bodin and Golden, 1981; Dondo and Cerdá, 2007; Özdamar and Demir, 2012; Schönberger, 2006; Simchi-Levi et al., 2005). Similarly, clustering has been used to understand the distribution of freight demand and simplify logistics operations (Cao and Glover, 2010; Sharman and Roorda, 2011; Singh et al., 2007; Qiong et al., 2011). However, these works present several limitations. In many cases revenues are not considered -or highly simplified- when demand bundles are constructed. Furthermore, uncertainty related to lane price and volume is not captured. On the other hand, clustering approaches used in the past focus on geographic proximity but cannot capture network effects resulting from the complex interdependencies among lanes. The main objective of this paper is proposing a systematic framework for demand clustering in freight logistics networks that overcomes these limitations. The contributions of the framework to literature are fourfold: (1) incorporating economic interdependencies among clustered lanes considering network effects, (2) considering market prices in the clustering process, (3) integrating uncertainty associated to variations on lane prices and volume, (4) developing a computationally efficient method. These contributions are demonstrated with numerical experiments.

The paper is organized as follows. Section 1 introduces and motivates this research. Section 2 reviews related literature. Section 3 clearly defines the problem to be solved. Section 4 presents the methodology to solve it. Section 5 presents numerical results and advantages. Section 6 summarizes the work and provides future research directions.

2. Literature review

This section reviews relevant literature related to carrier economies and network clustering. It is observed that an efficient method for demand clustering in freight logistics networks that accounts for shipment volume and price uncertainty is missing in literature. This motivates the development of the proposed model.

Finding groups of demand with synergetic properties in freight logistics networks is very important for strategic analysis, decision making, and business improvement at TL firms. However, detecting these lanes is not an easy task. Analyzing the exponential number of all the possible combinations of lanes (Song and Regan, 2003), prices and desired volumes is a hard combinatorial problem known as the lane bundling problem, where demand is grouped based on complementary characteristics. This intractable problem requires the computation of several NP-hard sub problems. Thus, bidding advisory models have been developed to study it in the context of TL CAs (Song and Regan, 2003, 2005; Wang and Xia, 2005; Lee et al., 2007; Chang, 2009; Huang and Xu, 2013; Xu and Huang, 2013, 2014). The underlying concept behind lane bundling is achieving economies of scope (Caplice, 1996; Jara-Diaz, 1983, 1981). Although recent bidding advisory models (Kuyzu et al., 2015; Triki et al., 2014; Ergun et al., 2007) represent competitors using stochastic prices, none of these works consider the uncertainty related to variations in freight flows and prices simultaneously.

Economies of scope are achieved by strategically positioning trucks such that follow-up loads are guaranteed and routing costs are distributed among several shipments. Backhauls are basic examples of economies of scope (Fig. 1). If a truck delivers a shipment from i to j with price p_{ij}^1, cost c_{ij}, and returns empty to i (cost c_{ji}), the expected profit will be $\Pi^1 = p_{ij}^1 - (c_{ij} + c_{ji})$. However, if there is a backhaul (loaded return) the profit is $\Pi^2 = p_{ij}^1 + p_{ji}^2 - (c_{ij} + c_{ji})$ where any price p_{ji}^2 increases profits ($\Pi^1 \leqslant \Pi^2$).

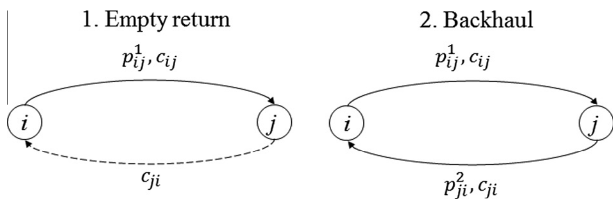

Fig. 1. Example of economies of scope.

In this work, the lane bundling problem is addressed using a clustering approach where subsets of elements sharing similar characteristics are grouped into clusters. In the last few years researchers and practitioners have used clustering methods to aggregate elements based on their proximity in multidimensional spaces, e.g., hierarchical, k-means, two-step, ad-hoc clustering, among others. Several vehicle routing problems (Bodin and Golden, 1981; Dondo and Cerdá, 2007; Özdamar and Demir, 2012; Schönberger, 2006; Simchi-Levi et al., 2005) take advantage of these methods by dividing the original network into subsets of geographically-close nodes where finding optimal routes is less cumbersome. Additionally, freight logistics problems have used clustering to understand the geographic distribution of demand and simplify logistics operations (Cao and Glover, 2010; Sharman and Roorda, 2011; Singh et al., 2007; Qiong et al., 2011). However, there are three limitations when proximity-based methods are used to cluster elements with an underlying network structure (Fortunato, 2010): (1) clustering points in a network requires at least a similarity metric for each pair of nodes, so storage space grows exponentially, (2) defining metric spaces to describe proximity in graphs is not trivial and significantly increases computational complexity, and (3) numerical experiments show that clusters highly depend on the type of metric defined.

Community detection algorithms (CDAs), e.g., Girvan and Newman (2002), Blondel et al. (2008), overcome these limitations. Refer to Fortunato (2010) for a comprehensive review on this topic. CDAs are developed to unmask highly interconnected elements in a network. Although they have been used to analyze several complex networks (e.g., social and biological networks, the World Wide Web, the international trade network), they are scarcely used in transportation applications. Nejad et al. (2012) is one of the few examples of using CDAs to understand transportation problems. Their work describes traffic conditions in highway networks but does not consider logistics operations or vehicle-routing. So, to the best of the authors' knowledge, community detection has neither been used in trucking research nor for the lane clustering problem. Nonetheless, CDAs are extremely important to consider network effects between lanes, i.e., economies of scope.

Applying CDAs in this context requires defining the elements to cluster and their level of interconnectivity. In this work these elements are lanes. For each pair of lanes the interconnectivity metric is defined as the utility of having them in the same cluster, i.e., served by the same trip-chain. Fan et al. (2006) also propose using utility functions to determine the proximity of clustered vehicles in vehicular ad-hoc networks (VANETS). They hypothesize utility functions based on available information. However, in this research utility is not explicitly available in the original transportation network (TN). Hence, a series of network transformations are required to construct an interconnectivity network (IN) suitable for community detection.

Demand clustering in freight logistic networks is important for the businesses and operations of carriers. Clear examples of such importance are the bidding advisor models developed to bundle services in TL combinatorial auctions. This review shows that efficient methods that account for uncertainty on shipment volume and price are needed. Likewise, the underlying network structure of this problem makes it suitable for clustering frameworks like CDAs. Nonetheless, this requires a proper definition of the network to cluster. These motivations encourage the development of an efficient and novel clustering framework that considers interdependencies between lanes and includes uncertainty related to lane volumes and prices. The specific problem solved in this research is defined in the next section.

3. Problem definition

This section describes the economic relationships in freight logistics networks served by TL carriers. Then the problem to be solved is clearly defined.

In general, the clients of TL companies are known as Shippers. Let a lane be the volume of truckloads per unit of time between an origin–destination (OD). Shippers are responsible for several lanes associated to their supply chains. They require transportation because they do not own transportation assets or because they own fleets but require additional capacity. TL carriers serve lanes of demand. A carrier can serve all or a subset of lanes for a specific shipper, and can work for many of them at the same time. TL companies operate over transportation networks (TNs). Their profits are determined by the right combination of prices and operational costs. Variable costs are related to loading/unloading activities, loaded, and empty movements. Clearly, TL carriers are only paid for loaded movements. So, minimizing empty trips by guaranteeing follow-up loads is vital for profitable operations. Deploying vehicles in places where little freight originates is undesirable. Although fixed costs impact firm finances, Nagle et al. (2011) suggest that it is sufficient to consider variable costs only when developing effective revenue management strategies. So, fixed costs are not considered in the analysis. Carriers explore economies of scope by strategically serving demand with the right balance between volume and topology.

Uncertainty affects the operation of businesses because forecasted demand and prices are used to cluster demand based on vehicle routing strategies. However, if the actual demand significantly differs from the forecasted one there are economic losses and discontent from the carrier, who might compensate by reducing its level of service. This, in turn, affects the regular operation of the shipper and its supply chain. A good understanding of demand uncertainty helps the carrier developing proper clusters of demand. A highly competitive environment forces TL carriers to choose market prices that are significantly interrelated to lane volumes. These elements are affected by common sources of uncertainty.

The problem solved by this research is clearly stated below. Table 1 summarizes mathematical notation. This paper considers a carrier serving a set of lanes $\widehat{D}$ and looking for the possibility of incorporating new lanes $D \setminus \widehat{D}$ into its logistics operation (D are all lanes considered in the problem). For each lane $i \in D$ historical observations of shipment prices $\mathcal{P}_i$ and lane volumes $\mathcal{Q}_i$ are available. They are organized in the $\mathscr{o} \times |D|$ matrices $\mathcal{P}$ and $\mathcal{Q}$ respectively, where $\mathscr{o}$ is the number of observations. The carrier operates over a TN $G(N, A)$, where, N are pickup/delivery nodes, and A are directed arcs connecting these nodes. Arcs $(o, d) \in A$ are associated to traversing costs c_{od}, i.e., for loaded or empty truck movements, and nodes $o, d \in N$ to pickup/delivery costs ς_o, ς_d. The carrier has a fleet of trucks of size $\mathscr{v}$. Given these characteristics of the carrier and TN, we are asked to find the clusters of demand $D^\ell, \ell = 1, \ldots, \mathcal{L}$ that represent increased expected profits for the carrier.

Table 1
Mathematical notation.

Notation	Definition
$G(N, A)$	Transportation network (TN) composed by a set of nodes N connected by the set of traversing arcs A
c_{od}	Traversing cost associated to each arc $(o, d) \in A$
D	Set of all lanes considered in the problem
$\widehat{D}$	Set of current lanes served by the carrier $\widehat{D} \subset D$
$\mathcal{L}$	Total number of clusters found by the algorithm
D^ℓ	ℓ^{th} cluster of lanes. $D^\ell \subset D, \ell = 1, \ldots, \mathcal{L}$
$F(\cdot \vert \mu, \sigma)$	Normal cumulative distribution function for mean μ and standard deviation σ
$f(o, d)$	Mapping from $o, d \in N$ to $i \in D$. $f : N^2 \to D$ such that demand in lane $i \in D$ is picked–up at $o \in N$ and delivered at $d \in N$
$\mathcal{G}(D, \mathcal{A})$	Demand super network composed by a set of demand nodes D connected by the set of traversing arcs $\mathcal{A}$
$g(i, j)$	Mapping from $i, j \in D$ to $d, o \in N$. $g : D^2 \to N^2$ such that $d \in N$ is the delivery node associated demand in lane $i \in D$ and $o \in N$ is the pickup node associated to demand in lane $j \in D$
$h(i)$	Mapping from $i \in D$ to $o, d \in N$. $g : D \to N^2$ such that demand in lane $i \in D$ is picked–up at $o \in N$ and delivered at $d \in N$
M	Number of samples selected for the Latin Hypercube Sampling process
$\mathscr{o}$	Numbers of historical observations of prices the corresponding shipment flows available to the carrier
P	$M \times \vert D \vert$ matrix of samples for each shipment price associated to lane $i \in D$
$\mathcal{P}$	$\mathscr{o} \times \vert D \vert$ matrix of observations for each shipment price associated to lane $i \in D$
$\bar{p}$	Vector of mean prices. $\bar{p} = \text{mean}(\mathcal{P})^T$
Q	$M \times \vert D \vert$ matrix of samples for each volume of shipments associated to lane $i \in D$
$\mathcal{Q}$	$\mathscr{o} \times \vert D \vert$ matrix of observations for each volume of shipments associated to lane volume $i \in D$
$\bar{q}$	Vector of mean volume of shipments. $\bar{q} = mean(\mathcal{Q})^T$
ς_i	Loading / unloading cost associated to serving lane $i \in D$
V	Covariance matrix for the observations $[\mathcal{P}\mathcal{Q}]$
$\mathscr{v}$	Number of available vehicles (fleet size)
$W(D, \omega)$	Demand super network composed by a set of demand nodes D and a set of undirected weighted links ω (interconnections)
x_{ij}	Flow of trucks repositioned to serve demand $j \in D$ after serving demand $i \in D$. $(i, j) \in \mathcal{A}$
χ_{od}	Flow of trucks traversing arc $(o, d) \in A$

4. Methodology

This section proposes an algorithmic approach to solve the problem formulated in Section 3, which is based on a series of network transformations illustrated in Fig. 2. Table 2 summarizes the pseudo code for the main algorithm which is supported by four modules.

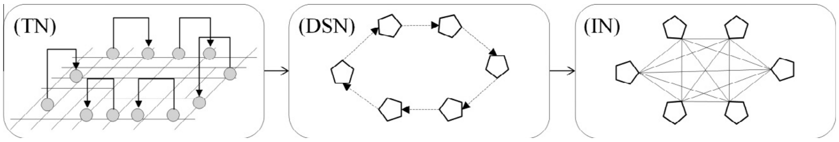

Fig. 2. Conceptual representation of network transformations.

Intuitively, the TN is composed by a set of nodes (pickup or delivery according to the lane distribution). Directed arcs between these nodes indicate traversing costs for loaded and empty trips (repositioned after delivering). Likewise, each shipment in a lane is associated to a price and pickup/delivery costs. Historical observations of prices and demand are used to design a number of scenarios according to their likelihood of occurrence and joint dependency. This is achieved using a Latin hypercube sampling method that accounts for dependency among sampled variables, i.e., price and demand level. Each sample determines an instance of prices and demand (truck volume) for the analyzed lanes. For each instance, a demand super network (DSN) – where nodes are lanes and directed arcs represent the repositioned flow of trucks between lanes – is constructed. A profit maximization linear program (LP) is used to find the optimal distribution of loaded and empty trips in the DSN. Each lane can be part of a trip-chain that connects several lanes and provides economies of scope to the carrier. However, there are two issues for proper demand clustering at this point: (1) flows are aggregated so it is not possible to differentiate trip-chain, and (2) – assuming trip-chains can be found – there is no evident connection between all lanes in a trip-chain (only the downstream and upstream connections are known). So, a novel method is proposed to detect and disaggregate trip-chains, i.e., tours composed by synergetic lanes in the DSN. The joint utility between every pair of demand in these tours is computed and used to generate an interconnectivity network (IN) where each pair of lanes is weighted using the bilateral utility of having them in the same tour. This network is updated after running each sampled scenario. Then, when all scenarios are explored, a CDA is applied over the IN taking advantage of the rich information accumulated by the sampling process and revealing the corresponding clusters of profitable demand.

Table 2
Main algorithm: Demand clustering in freight logistics networks.

Step	Description
1	$\bar{p}, \bar{q}, V \leftarrow \text{mean}(\mathcal{P})^T, \text{mean}(\mathcal{Q})^T, \text{cov}([\mathcal{P}\mathcal{Q}])$
2	$[PQ] \leftarrow$ **latinHypercubeSampleNormal**$\left(\begin{bmatrix}\bar{p}\\ \bar{q}\end{bmatrix}, V, M\right)$ **Module 1**
3	$\omega \leftarrow \lvert D\rvert \times \lvert D\rvert$ matrix: $\omega_{ij} = 0$
4	For $m = 1, \ldots, M$
5	$p^T, q^T \leftarrow m$ th row of P, mth row of Q
6	$x^m \leftarrow$ **demSupNetLP**$(c_{od}, \varsigma_i, \widehat{D}, p, q, \nu)$ **Module 2**
7	$\omega \leftarrow \omega +$ **updateInterconnections**$(x^m, p, c_{od}, \varsigma_i)$ **Module 3**
8	End
9	If $(\omega_{ij} < 0)$
10	$\omega_{ij} \leftarrow 0$
11	Else
12	$\omega_{ij} \leftarrow \frac{\omega_{ij}}{m}$
13	End
14	$D^1, \ldots, D^{\mathcal{L}} \leftarrow$ **clustering**(ω) **Module 4**
15	Return $D^1, \ldots, D^{\mathcal{L}}$

In general, problems that are affected by uncertainty are solved using stochastic programming, robust optimization, or scenario analysis (e.g., Ma et al., 2010; Patil and Ukkusuri, 2007; Ukkusuri, 2005; Remli and Rekik, 2013). Although the solutions of stochastic programs present the right balance between risks and benefits, the tractability of these methods is significantly affected by the curse of dimensionality. Furthermore, the combinatorial nature of the clustering problem aggravates this limitation. On the other hand, problems solved using robust optimization are more tractable. Nonetheless, their low-risk solutions significantly underestimate benefits. This research follows a scenario analysis approach where uncertainty is addressed by sequentially testing several interdependent scenarios generated from the sampling process. As solutions are optimized for each DSN instance, the IN is updated with the bilateral utility of having two lanes in the same trip-chain. If a demand duplet appears in the optimal solution for several scenarios, the strength of the corresponding utility is reinforced. After testing all interdependent scenarios, the CDA uses information accumulated through this process to construct the clusters. This balances the pros and cons of stochastic and robust optimization. It is computationally tractable and relaxes conservative solutions by fathoming several instances of the problem. Thus, the resulting clusters balance profitability under different realizations of demand and price, which adds a desirable level of risk without sacrificing tractability.

Formally, the algorithm starts by computing the mean $\bar{p}, \bar{q}$ and covariance V of historical observations $\mathcal{P}$ and $\mathcal{Q}$ to generate M dependent samples from a Latin Hypercube sampling process, i.e., $P \in \mathbb{R}^{M\times|D|}$ and $Q \in \mathbb{R}^{M\times|D|}$ (Module 1). A sufficiently large number of samples M is defined by the modeler. For each sample $m \in \{1,\ldots,M\}$ an instance of DSN is generated and a profit maximization network flow LP is solved to find the optimal distribution of trucks x^m that maximizes carriers profits (Module 2). Then, each resulting trip-chain is fathomed to determine the utility between duplets of lanes ω_{ij} and update the IN (Module 3). After properly standardizing ω_{ij}, a CDA is used to unmask the demand clusters D^ℓ (Module 4).

4.1. Module 1: Latin hypercube sampling with dependent variables

A sampling process is used to replicate stochastic demand and prices. Sampling is a common technique in experiment design and scenario testing. The Monte Carlo method (Metropolis and Ulam, 1949) is a popular procedure but it is expected to generate biased samples. The Latin hypercube sampling (McKay et al., 1979; Iman et al., 1981) overcomes this limitation by evenly distributing the multidimensional space (Latin hypercube) and selecting samples from each subdivision. However, this approach cannot capture flow and price dependency which is important as trucking volumes and prices are not independent. For example, fluctuations in the flow of trucks delivering the final demand of a product proportionally affect the movement of goods in the upstream supply chain. Similarly, economies of scope correlate prices and volumes, e.g., high volume of truckloads in one direction and low volume in the opposite one might result in lower prices for the backhauls. Stein (1987) proposes a variation of the Latin hypercube sampling that considers dependency between variables. Therefore, that method is used in this module.

Latin hypercube sampling with variable interdependency is a useful tool when limited data about volume and price fluctuations are available. In the worst case, estimations of the mean and covariance are required. Although this work assumes normal distributions to represent these variations, virtually any type of distribution can be assumed if it is properly supported. This is one of the benefits of developing a modular clustering framework.

Table 3 summarizes the pseudo code for this module. The vector of average values $\mu = \left[\bar{p}^T \bar{q}^T\right]^T$ and the corresponding covariance matrix V are used to generate M samples from a multivariate normal distribution $z \in \mathbb{R}^{M\times|\mu|}$. These values are ranked column-wise to divide the space into M independent subdivisions, which are standardized in the interval $[0,1]$ and assigned to the middle of each range $\phi \in \mathbb{R}^{M\times|\mu|}$. Finally, the matrix of samples $y = [PQ] \in \mathbb{R}^{M\times|\mu|}$ is populated using the values y_{ij} for which the normal cumulative distribution function $F(y_{ij}|\mu_j, \sqrt{V_{jj}})$ is equivalent to ϕ_{ij}.

Table 3
Module 1: Latin hypercube sampling with dependent variables.

Step	Description
1.1	$z \leftarrow M \times \lvert\mu\rvert$ matrix where each row is a sample with multivariate normal distribution (μ, V)
1.2	$\phi \leftarrow M \times \lvert\mu\rvert$ matrix where ϕ_{ij} correspond to the ranking of z_{ij} in the jth column of z
1.3	$\phi \leftarrow (\phi - 0.5)/M$
1.4	$y \leftarrow m \times \lvert\mu\rvert$ matrix where y_{ij} corresponds to: $y_{ij} \leftarrow y_{ij} : \phi_{ij} = F\left(y_{ij}\lvert\mu_j, \sqrt{V_{jj}}\right) = \frac{1}{\sqrt{2\pi V_{jj}}}\int_{\infty}^{y_{ij}} e^{\frac{-(t-\mu)^2}{2V_{jj}}} dt$
1.5	Return y

4.2. Module 2: Demand super network linear program

This module constructs the DSN first and then solves a network-flow LP to find the flow of trucks that maximizes profits in this network.

The DSN is described as follows. Let $\mathcal{G}(D, \mathcal{A})$ be the DSN where the set of super nodes corresponds to the set of lanes D. Nodes in D are connected by a set of directed arcs $\mathcal{A}$, where $(i,j) \in \mathcal{A}$ represents the trucks repositioned to serve demand $j \in D$ after serving demand $i \in D$. The following network transformations are illustrated in Fig. 3. Each arc is associated with a repositioning utility defined as $u_{ij} = p_j - c_{g(i,j)} - c_{h(j)} - \varsigma_j$, where p_j is the current sampled price and ς_j is the loading/unloading costs for lane $j \in D$, $c_{g(i,j)} = c_{d_i o_j}$ is the traversing cost of a truck repositioned from $d_i \in N$ (Node where demand $i \in D$ is delivered) to $o_j \in N$ (Node where demand $j \in D$ is picked up), and $c_{h(j)} = c_{o_j d_j}$ is the traversing cost of a truck serving the

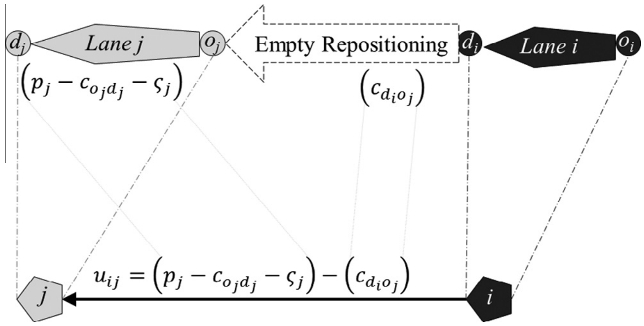

Fig. 3. Arc representation in the DSN and its relationship with the TN.

downstream demand $j \in D$ picked up at $o_j \in N$ and delivered at $d_j \in N$. The mapping functions $g(i,j) = (d_i, o_j) \in A$ and $h(j) = (o_j, d_j) \in A$ are conveniently defined to make transformations between $\mathcal{G}(D, \mathcal{A})$ and $G(N, A)$.

Subsequently, the LP (1)–(6) is solved. Here, the variables x_{ij} represent the flow of repositioned trucks. The Objective Function (1) maximizes the utility associated to the deployment of x over $\mathcal{G}(D, \mathcal{A})$ such that Constraints (2)–(6) are satisfied. Constraint (2) indicates flow conservation for the trucks serving each lane $j \in D$. Constraint (3) states that demand in the set of lanes currently served by the carrier $\widehat{D}$ most be served. On the other hand, Constraint (4) specifies that lanes in the set of potential demand to be included in the carrier network $D \setminus \widehat{D}$ are optionally served. Constraint (5) designates that there is a limited availability of trucks υ to serve the network. Finally, Constraint (6) stipulates the non-negativity nature of x_{ij}. Notice that this LP can efficiently be solved by regular commercial software, e.g., CPLEX.

$$\max \sum_{(i,j) \in \mathcal{A}} u_{ij} x_{ij} \tag{1}$$

s.t.

$$\sum_{i \in D} x_{ij} = \sum_{i \in D} x_{ji}, \quad \forall j \in D \tag{2}$$

$$\sum_{i \in D} x_{ij} = q_j, \quad \forall j \in \widehat{D} \tag{3}$$

$$\sum_{i \in D} x_{ij} \leqslant q_j, \quad \forall j \in D \setminus \widehat{D} \tag{4}$$

$$\sum_{i \in D} x_{ij} \leqslant \upsilon, \quad \forall j \in D \tag{5}$$

$$x_{ij} \geqslant 0, \quad \forall (i,j) \in \mathcal{A} \tag{6}$$

4.3. Module 3: Update interconnections

This module finds each tour in the network and relates each duplet of demand $i, j \in D$ with a weight ω_{ij} in the IN. The pseudo code presented in Table 4 describes this process. First each flow x_{ij} in the DSN is associated with the corresponding flows in the TN, i.e. $\chi_{g(i,j)} = \chi_{d_i o_j}$ and $\chi_{h(i)} = \chi_{o_i d_i}$. Then arcs A in the TN are locally modified to consider only arcs with flow. The main loop searches trip-chains in the network. At each iteration, the arc $(s, r) \in A$ with less flow χ_{sr} is selected and removed from A. Then, the shortest path T from r to s is computed. Its cost is $c(T)$. Each flow χ_{ij} associated to arcs in T, and arc (s, r) itself is reduced by χ_{sr}. Subsequently a set of lanes $\mathcal{T}$ is generated to hold the demand elements associated to $T \cup \{(s, r)\}$. Notice that the mapping function $f(o_i, d_i) = i \in D$ is used to map elements from $G(N, A)$ to $\mathcal{G}(D, \mathcal{A})$. Then, the average cost associated to each element in $\mathcal{T}$ is computed and the interconnectivity between elements in each tour is updated by adding the fractional income associated to the demand objects i and j minus the corresponding average cost.

Table 4
Module 3: Update interconnections.

Step	Description
3.1	$\chi \leftarrow \lvert N\rvert \times \lvert N\rvert$ matrix
3.2	$\chi_{g(i,j)} \leftarrow \chi_{g(i,j)} + x_{ij} \forall (i,j) \in \mathcal{A}$
3.3	$\chi_{h(i)} \leftarrow \chi_{h(i)} + x_{ij} \forall (i,j) \in \mathcal{A}$
3.4	$A \leftarrow \varnothing$
3.5	$A \leftarrow A \cup \{(o,d), \forall o,d \in N : x_{od} > 0\}$
3.6	$\omega \leftarrow \lvert D\rvert \times \lvert D\rvert$ matrix: $\omega_{ij} = 0$
3.7	While $(\max(x) > 0)$
3.8	$(s,r) \leftarrow argmin(\chi_{od} : (o,d) \in A)$
3.9	$A \leftarrow A \setminus \{(s,r)\}$
3.10	$T, c(T) \leftarrow$ compute shortest path from $r \in N$ to $s \in N$ over $G(N,A)$ using cost matrix c. Return path $T \subset A$ and its corresponding cost $c(T)$
3.11	$\chi_{od} \leftarrow \chi_{od} - \chi_{sr}, \forall (o,d) \in T \cup \{(s,r)\}$
3.12	$\mathcal{T} \leftarrow \varnothing$
3.13	$\mathcal{T} \leftarrow \mathcal{T} \cup \{f(o,d) \in D, \forall (o,d) \in T \cup \{(s,r)\}\}$
3.14	$\bar{c}(\mathcal{T}) \leftarrow \chi_{sr} \frac{c(T)+c_{sr}}{\lvert\mathcal{T}\rvert}$
3.15	If $(\lvert\mathcal{T}\rvert = 1)$
3.16	$\forall i \in \mathcal{T} : \chi_{sr}(p_i - \varsigma_i) - \bar{c}(\mathcal{T}) > 0$ $\omega_{ii} \leftarrow \omega_{ii} + \chi_{sr}(p_i - \varsigma_i) - \bar{c}(\mathcal{T})$
3.17	Else
3.18	$\forall i,j \in \mathcal{T} : i < j, \chi_{sr} \frac{(p_i+p_j)-(\varsigma_i+\varsigma_j)}{\lvert\mathcal{T}\rvert-1} - \bar{c}(\mathcal{T}) > 0$ $\omega_{ij} \leftarrow \omega_{ij} + \chi_{sr} \frac{(p_i+p_j)-(\varsigma_i+\varsigma_j)}{\lvert\mathcal{T}\rvert-1} - \bar{c}(\mathcal{T})$
3.19	End
3.20	End
3.21	Return ω

4.4. Module 4: Clustering

Module 4 (described by the pseudo code in Table 5) applies the community detection algorithm presented in Blondel et al. (2008). This algorithm is based on modularity maximization. Modularity "measures the density of links inside communities as compared to links between communities" (Blondel et al., 2008). This method has been successfully and efficiently used to detect network clusters in several applications. Fortunato's (2010) review highlights the multiple advantages of this algorithm. It can be used to analyze weighted directed networks, which is not the case for several efficient algorithms in literature, e.g., Girvan and Newman, 2002. Numerical experiments show that the algorithm is extremely fast and tractable for graphs with up to 10^9 edges, the modularity maxima found by the method are better than other greedy techniques, e.g., Clauset et al. (2004), and computational times outperform other modularity-based methods. Some limitations of this algorithm include the possibility of finding spurious partitions and variability in the clusters based on the order in which nodes are considered.

Table 5
Module 4: Clustering.

Step	Description
4.1	$\omega \leftarrow \omega + \omega^T$
4.2	$D^i \leftarrow \{i\}, \forall i \in D$
4.3	$\tau \leftarrow 0$
4.4	$\Theta_\tau \leftarrow 0$
4.5	$\Theta_{\tau+1}, D^\ell \leftarrow$ **computeModularity** (ω, D^i) **Sub-module 5**
4.6	While $(\Theta_{\tau+1} > \Theta_\tau)$
4.7	$\tau \leftarrow \tau + 1$
4.8	$D \leftarrow \{i : D^i \neq \varnothing\}$
4.9	$\omega \leftarrow \lvert D\rvert \times \lvert D\rvert$ matrix. $\omega_{ij} = 0$
4.10	$\omega_{ij} \leftarrow$ Weight of links between D^i and D^j
4.11	$D^i \leftarrow \{i\}, \forall i \in D$
4.12	$\Theta_{\tau+1}, \mathrm{D}^\ell \leftarrow$ **computeModularity**(ω, D^i) **Sub-module 5**
4.13	End
4.14	Return $D^1, \ldots, D^\ell, \ldots$

The main input for this algorithm is the interconnectivity matrix ω, which is first added to its transpose to standardize directed weights to the undirected case. The algorithm starts assigning each demand i to a cluster D^i. Then, initial clusters are recomputed based on modularity maximization sub-module (Sub-module 5). Next, the main while loop runs and sequentially aggregates clusters up to finding the configuration with the maximum modularity.

Since carriers are interested in detecting new clusters inside previously found clusters, for every cluster D^ℓ Module 4 is recursively applied. Thus, the initial clusters are defined as mega-clusters (MC). Each MC is composed by several interior sub clusters (SC). Consecutively, interior SCs are composed by smaller SCs and so on. This hierarchical clustering groups lanes in several strata.

4.4.1. Sub-module 5: Compute modularity

This sub-module (Table 6), which is also described in Blondel et al. (2008), iteratively swaps nodes between clusters. When there is increment in modularity $\Delta\Theta'$ by adding a node i to a cluster D^{λ} this action is performed. The process stops when modularity cannot be increased. Although this is a greedy approach, it has shown to be very efficient in practical settings.

Table 6
Sub-module 5: Compute modularity.

Step	Description
5.1	$g \leftarrow 1$
5.2	While $(g = 1)$
5.3	$g = 0$
5.4	For $i = 1, \ldots, \lvert D \rvert$
5.5	$\Delta\Theta \leftarrow 0$
5.6	$\lambda \leftarrow \{\lambda \in D : i \in D^{\lambda}\}$
5.7	For $j = 1, \ldots, \lvert D \rvert : \omega_{ij} > 0, j \in D^{\ell}, D^{\ell} \cap \{i\} = \emptyset$
5.8	$\dot{k}_{rs} \leftarrow \sum_{r,s \in D^{\ell}} (\omega_{rs})$
5.9	$\breve{k}_{is} \leftarrow \sum_{s \in D^{\ell}} (\omega_{is})$
5.10	$K \leftarrow 1/2 \sum_{\imath,\jmath,\in D} (\omega_{\imath\jmath})$
	$k_{\imath s} \leftarrow \sum_{\imath \in D} \sum_{s \in D^{\ell}} (\omega_{\imath s})$
5.11	$\hat{k}_{\imath i} \leftarrow \sum_{\imath \in D} (\omega_{\imath i})$
5.12	$\Delta\Theta' \leftarrow \left[\frac{\dot{k}_{rs} + 2\breve{k}_{is}}{2K} + \left(\frac{k_s + \hat{k}_i}{2K}\right)^2\right] - \left[\frac{\dot{k}_{rs}}{2K} - \left(\frac{k_s}{2K}\right)^2 - \left(\frac{\hat{k}_i}{2K}\right)^2\right]$
5.13	If $(\Delta\Theta < \Delta\Theta')$
5.14	$\Delta\Theta \leftarrow \Delta\Theta'$
5.15	$\lambda \leftarrow \ell$
5.16	$g = 1$
5.17	End
5.18	End
5.19	End
5.20	End
5.21	$\Theta_{\tau+1} \leftarrow \frac{1}{2K} \sum_{i,j \in D} \left(\omega_{ij} - \frac{\hat{k}_{\imath i}\hat{k}_{\imath j}}{2K}\right) \delta(i,j), \delta(i,j) = \begin{cases} 1, & \text{if } i,j \in D^{\ell} \\ 0 & \text{otherwise} \end{cases}$
5.22	Return $\Theta_{\tau+1}, D^{\ell}$

In summary, clusters of lanes are found using interdependent historical information for volume and price on every lane. Latin-hypercube is used to sample dependent volume/price scenarios. The optimal distribution of flow between lanes is determined for each sample solving a profit maximization LP. Synergetic lanes are interconnected based on their bilateral utility generating an interconnectivity network that is updated iteratively. Finally, community detection is used to cluster the network that emerges and finding profitable demand collections. An important benefit of this method is its flexibility to be implemented in well-known programming platforms like Matlab, Python, C++, Java, among other. Furthermore, each module can be either developed or borrowed from available open sources or commercial software. For example, Latin hypercube sampling is available in platforms like Matlab, R, Python, SAS/JMP, etc. Linear programing can be solved using commercial software, e.g., AMPL/CPLEX, ILOG CPLEX, Gurobi, Lindo, Gams, Matlab, etc. Source code for community detection algorithms is available for Matlab, C++, Python, among other, and implemented in several network analysis software, e.g., NetworkC and Gephi.

5. Numerical results

This section presents 2 theoretical numerical examples to illustrate the methodological framework. The first small example is used to visualize the performance of the method. Since the simplicity of this example cannot show the full potential of the method, a larger example is presented. Afterwards, a numerical experiment is performed to test the scalability of the proposed demand clustering methodology. The suite of algorithms is coded in Matlab and run in an average desktop with Inter ® Core 2 Duo Processor Processor (E8400) at 3.00 GHz and 4.00 GB of RAM. The open source code developed by Scherrer and Blondel (2014) is used for community detection.

The first numerical example is based on the TN in Fig. 4. Two lanes are included in the set of lanes considered for new businesses, i.e., $D \setminus \widehat{D} = \{(e,f),(c,d)\} = \{1,2\}$. Likewise, the carrier is currently serving one lane, i.e., $\widehat{D} = \{(a,b)\} = \{3\}$. Sufficiently large truck capacity is assumed. The total costs related to each lane (traversing plus loading/unloading) are presented in Table 7. Likewise, this table shows the corresponding repositioning costs after delivering. It is less expensive to reposition trucks after delivering in the new Lane 1 to the pickup demand at the current Lane 3 and vice versa. Repositioning from these two lanes to the new Lane 2 is more expensive.

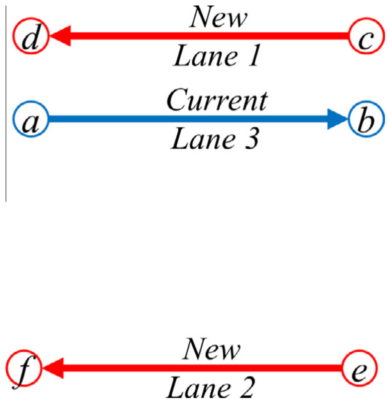

Fig. 4. Numerical example 1: Transportation Network (TN).

Table 7
Example 1: Costs in the TN.

From	To	Total Cost	Type
c	d	4	New Lane 1
e	f	4	New Lane 2
a	b	4	Current Lane 3
b	a	4	Repositioning Arc
b	c	1	Repositioning Arc
b	e	3	Repositioning Arc
d	a	1	Repositioning Arc
d	c	4	Repositioning Arc
d	e	8	Repositioning Arc
f	a	3	Repositioning Arc
f	c	8	Repositioning Arc
f	e	4	Repositioning Arc

Mean values for prices $\bar{p}$ and levels of demand $\bar{q}$ are available in Table 8. The corresponding covariance matrix V within and between price and demand level for each pair of lanes is available in Table 9, where negative values indicate opposite behavior between the observations. These are the main inputs to collects samples from Module 1. Mean prices for the new Lane 2 are in average higher than those from Lanes 1 and 3. Likewise, the levels of demand are in average higher for Lane 2 and lower for Lane 1. The covariance matrix shows that demand and prices have higher variability for Lane 2. In general, demand and prices have opposite trends, i.e., the more the demand the lower the price. Finally, demand and prices in Lane 2 have opposite trends to the values for Lanes 1 and 3.

Table 8
Example 1: Mean levels of demand and prices for each lane.

Lanes D	Mean level of demand $\bar{q}$	Mean price $\bar{p}$
1	80	10
2	200	30
3	100	10

Table 9
Example 1: Covariance for levels of demand and prices between lanes.

V		Demand			Price		
		1	2	3	1	2	3
Demand	1	400	−2000	400	−20	300	−40
	2	−2000	10,000	−2000	100	−1500	200
	3	400	−2000	400	−20	300	−40
Price	1	−20	100	−20	1	−15	2
	2	300	−1500	300	−15	225	−30
	3	−40	200	−40	2	−30	4

Three samples (i, ii, iii), i.e., $M = 3$, are used to illustrate the application of the model. The sampled values of price P and demand Q are presented in Table 10. These samples are collected from the Latin Hypercube Sampling technique presented in Module 1 based on the values for $\bar{p}, \bar{q}$, and V. They follow the trends observed on the dataset and discussed before.

Sampled prices P and levels of demand Q are used to construct instances of the optimization problem in Module 2. For each of these instances, Table 11 presents the utilities u_{ij} between lanes. Although high prices on Lane 2 increase the related marginal utilities for earlier instances, these values tend to decrease and utilities for other lanes increase.

Each instance is optimized and Table 12 shows the corresponding optimal values for the objective function. Instance *ii* presents the best system profits followed by *i* and *iii*. Notice that this variation is related to the uncertainty addressed in this paper.

Furthermore, Table 13 shows truck flows x_{ij} related to each optimized instance. Lanes 1 and 3 tend to complement each other. On the other hand, Lane 2 seems to perform better by serving itself (backhaul). In optimal conditions, a small amount of flow is repositioned between the current Lane 3 and the new Lane 2. In optimal conditions it would be profitable serving the 3 lanes together. However this is not always possible when carriers are asked to prioritize (cluster) subsets that would be more desirable.

Therefore, interconnections between lanes ω_{ij} are estimated (Module 3) and updated after solving each instance of the optimization problem. Then, the corresponding matrix is standardized as described on the Main Algorithm. The corresponding interconnectivity matrices are presented in Table 14. These results reinforce the importance of serving Lane 2 using backhauls and also the strong synergy that exists between Lanes 1 and 3.

Table 10
Example 1: Latin Hypercube Sampling.

	Samples	Lanes D 1	2	3
Demand Q	*i*	100	100	120
	ii	80	200	100
	iii	60	300	80
Price P	*i*	9	45	8
	ii	10	30	10
	iii	11	15	12

Table 11
Example 1: Utility between lanes related to each sample.

	Instance *i*			Instance *ii*			Instance *iii*		
Lanes	1	2	3	1	2	3	1	2	3
1	1	33	3	2	18	5	3	3	7
2	−3	37	1	−2	22	3	−1	7	5
3	4	38	0	5	23	2	6	8	4

Table 12
Example 1: Objective function values for each sample.

Instance	*i*	*ii*	*iii*
Objective function	4440	5280	3000

Table 13
Example 1: Truck flow repositioned between lanes for each sample.

	Instance *i*			Instance *ii*			Instance *iii*		
Lanes	1	2	3	1	2	3	1	2	3
1	0	0	100	0	0	80	0	0	60
2	0	80	20	0	180	20	0	280	20
3	100	20	0	80	20	0	60	20	0

Table 14
Example 1: Intermediate and standardized temporal interconnectivities.

	Instance *i*			Instance *ii*			Instance *iii*			Standardized		
Lanes	1	2	3	1	2	3	1	2	3	1	2	3
1	0	0	200	0	0	600	0	0	1080	0	0	360
2	0	2640	660	0	5880	1060	0	6720	1200	0	2240	400
3	200	660	0	600	1060	0	1080	1200	0	360	400	0

Finally, the standardized interconnections are used to construct the demand clusters $D^{\mathcal{L}}$ as described in Module 4. These results are presented in Table 15 which shows that, if the carrier is asked to cluster complimentary lanes, it is better to group Lanes 1 and 3 (D^1) and Lane 2 performs well served with backhauls (D^2).

Table 15
Example 1: Demand clusters.

Cluster 1 (D^1)	Cluster 2 (D^2)	Modularity
{1,3}	{2}	0.21

In an ideal situation the carrier can configure its network such that all demand is served optimally. However, there are cases when it has to cluster or prioritize subsets of demand to be served conjointly under uncertain prices and levels of demand, e.g., combinatorial auctions, new business opportunities, among others. This method provides good quality clusters that enhance the profits related to synergetic lanes served together. Thus, Table 16 compares the profits expected from each cluster with respect to the system optimal, which gives a sense of its quality. Clearly, the sum of profits for independent clusters is very similar to the system optimal (gap below 1.3%).

Table 16
Example 1: Profit comparison.

Instance	Optimal profit				Gap
	D^1	D^2	$D^1 + D^2$	D	
i	720	3700	4420	4440	0.5%
ii	840	4400	5240	5280	0.8%
iii	860	2100	2960	3000	1.3%

Given the simplicity of the previous example, a larger theoretical problem is presented to illustrate the value of this method. The second numerical example is based on the TN in Fig. 5(a). Each arc in the grid network has unitary cost. Without loss of generality, it is assumed that the cost for each lane (traversing plus loading/unloading) is equivalent to the sum of unitary costs for covered arcs. Repositioning costs correspond to the shortest path between lanes in the grid network. Currently, the carrier serves $|\widehat{D}| = 21$ lanes and is considering other 21 lanes for new businesses. In total, this analysis considers $|D| = 42$ lanes. A number of $\wp = 100$ contemporaneous observations for price $\mathcal{P}$ and shipment volume $\mathcal{Q}$ are available for each lane. The mean $[\bar{p}, \bar{q}]$ and covariance V for these values are illustrated in Fig. 5(b).

The carrier selects $M = 100$ samples to undertake the analysis (Module 1). For each sample, the linear program in Module 2 is solved and the IN populated (Module 3). Fig. 6 presents the resulting IN and shows that several lanes present synergies when operated together. However, these synergies are stronger for groups of them. For example, the new lane 7 is strongly related to the current lane 22, which is intuitive by the directionality of the flows in Fig. 5. Furthermore, current lanes 30 and 32 complement these movements by reducing empties. Notice that the geographic position of 30 and 32 results in no direct interconnection between them but they have strong common allies, i.e., 7 and 22. Similarly, the new lane 15 forms a strong

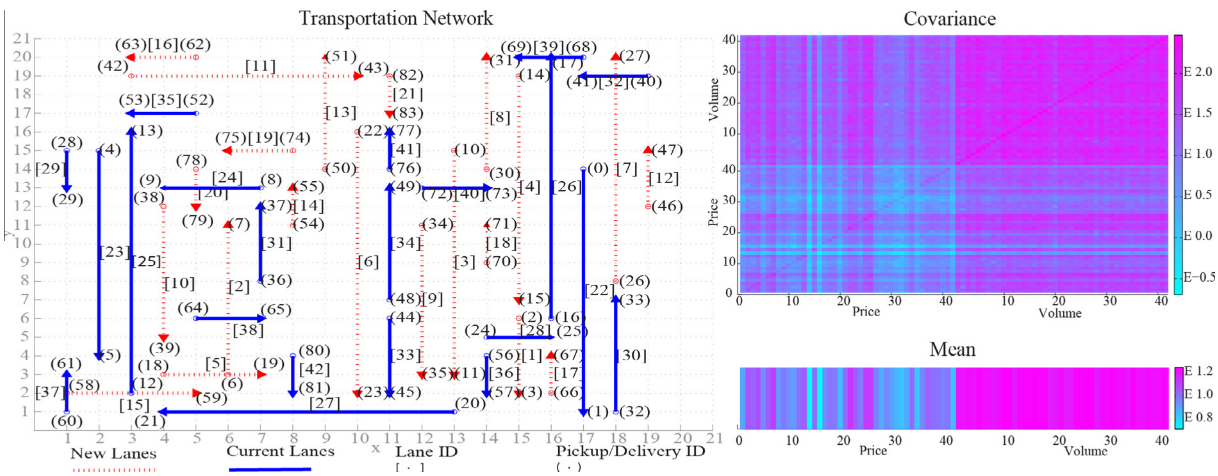

Fig. 5. Numerical example 2: (a) TN and demand (left), (b) mean and covariance for price and truck volumes (right).

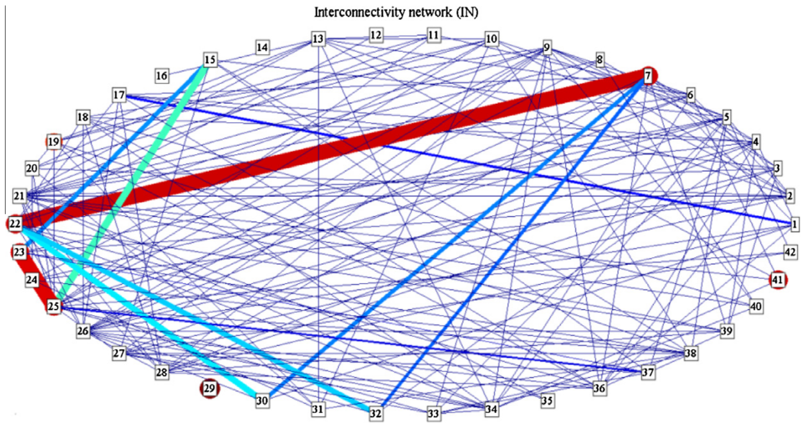

Fig. 6. Numerical example 2: Interconnectivity Network (IN).

triplet with lanes 23 and 25 giving continuity to the current traffic flows. On the other hand, there are isolated lanes with scarce interconnections but strong connectivity to themselves, i.e., new lane 19 and current lanes 29, 35, 41. These lanes are characterized by backhaul movements and this can happen for several reasons, e.g., they are isolated or peripheral in the network, the topological characteristics of lanes in their neighborhoods are not suitable for follow-up loads, neighbor lanes have stronger synergy with other lanes in the system. Interestingly, lane 29 has no interconnections but its self-strength is extremely high, i.e., it has no synergy but is very valuable for the carrier. This is because it is a profitable but peripheral lane. Other groups of lanes hidden in the IN are mined using community detection (Module 4).

The clustering algorithm reveals seven MCs (Fig. 7(a)). Community detection reinforces the intuition presented above by unmasking synergies not distinguishable by observation. 22 MCs are observed, i.e., 7 aggregating more than two lanes and 15 are singletons. MC 1 is composed by lanes 7, 22, 30, 32 as noticed above. Synergies are complemented by the new lanes 4, 8, 3 and current lanes 26, 39. MC 2 is composed by lanes 15, 23, 25 -noticed before- and complemented with the current lane 37. Other clusters are MC 3 composed by new lanes 18, 9, 6, current lane 34, MC 4 by new lanes 11, 13, 16 only, MC 5 by new lanes 5, 2, current lane 27, by new lanes 1, 17 only, MC 7 by new lane 20, current lane 24. Each of the remaining lanes is a cluster itself. Lanes 19, 29, 35, 41, mentioned above, are in this category. Interestingly, many current lanes are benefited by adding new lanes. On the other hand, clusters composed only by new lanes represent new business opportunities for the carrier.

The hierarchical structure of the clusters is obtained by fathoming MCs. Fig. 7(b) shows the composition of the MCs and their corresponding SCs. MC 1 is divided in two SCs: SC 1.1 with strong interconnected lanes and SC 1.2 with other interconnected lanes that have less strength, MC 2 segregates lane 37 and creates SC 2.1 with the strong triad 15, 23, 25. Furthermore, lanes 18 and 5 are separated from M3 and M5 creating new SCs. MCs 4, 6, 7 are strong by themselves and no disaggregation is needed. This example shows that analysing the freight demand clustering problem is considerably complex even for small instances. The proposed methodology reduces this complexity and is a viable alternative for carriers that face large instances of this problem in their regular operations.

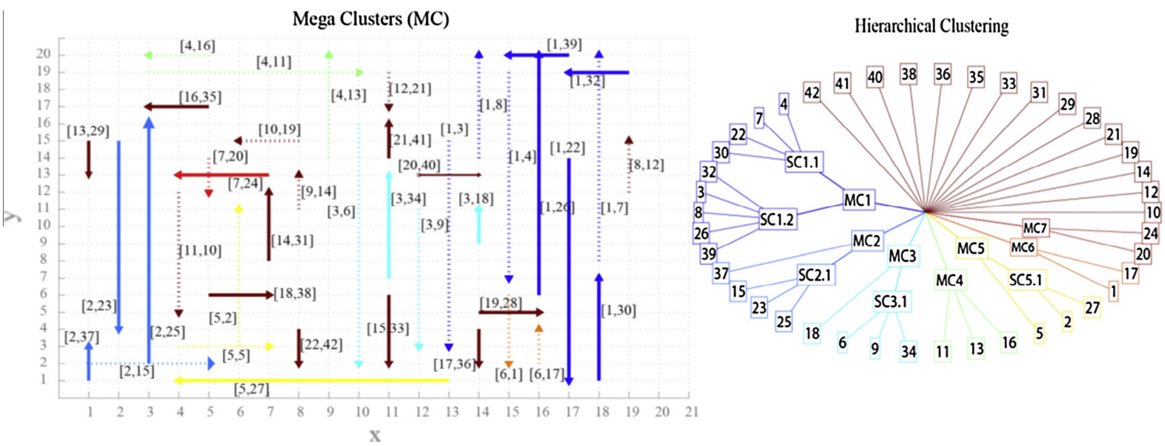

Fig. 7. Numerical Example 2: (a) MCs of demand (notation: [MC ID, lane ID]) (left), (b) hierarchical clustering (right).

The scalability of the method is tested with a numerical experiment. The number of samples in the experiment is set to $M = 100$. The geography of the transportation network is randomly generated with traversing cost equal to the Euclidean distance between nodes. Likewise, the set of lanes D and the corresponding sets of observation P and Q are synthetically generated following appropriate ranges avoiding inconsistencies. Table 17 summarizes the experiment where demand varies from 25 to 500 lanes and the corresponding pickup/delivery nodes go from 50 to 1000.

Table 17 shows that the method is suitable for sufficiently large instances. The modules that are spending the most computational time are the one related to the solution of the LP (Module 2) and the one where trip-chains are searched to update the IN (Module3). Likewise, modularity and number of clusters increases as the number of demand objects increases. In general, the number of MCs (computed before starting the recursive process described in Module 4) represents a large proportion of the total clusters found.

Table 17
Scalability experiments.

Demand	Nodes	MCs	Modularity	Total clusters	CPU Time (s)					
					Inputs	Module 1	Module 2	Module 3	Module 4	Total
25	50	6	0.71	12	0.00	0.00	2.76	1.56	1.08	5.40
50	100	21	0.83	32	0.11	0.02	8.44	7.46	1.19	17.21
100	200	30	0.83	47	0.03	0.05	30.09	40.06	5.51	75.74
200	400	108	0.83	135	0.47	0.31	189.17	233.03	19.00	441.98
500	1000	284	0.90	340	1.22	4.06	2836.70	3212.40	28.83	6083.20

There are several key insights from these results. Network effects must be considered when clustering freight demand. Although geographic proximity highly impacts clustering, it is not the only and most important attribute. Bilateral utility between lanes determines their actual proximity, which is a function of the trip-chains encompassing them. Thus, topology (geography and directionality), shared profits (volumes, costs, and prices), and contemporaneity, are key elements for demand clustering in freight logistics networks affected by uncertainty. High bilateral utility is a key trait for clustering demand but it is not sufficient. The strength and degree of interconnectivities between lanes determine their actual closeness, in social networks jargon: "the friend of your friend is likely also to be your friend" (Newman, 2003). Furthermore, lanes complement at different levels. Those with higher synergies remain together over several sub-clusters. Lanes with less strength either disconnect leaving the stronger elements clustered, or agglomerate into new sub clusters with other synergetic lanes. Not all lanes are synergetic in the system. Some of them are not suitable to be clustered and they operate better alone. This happens because they are distant, i.e., geographically far, with opposite directionalities, or not competitive with respect to other lanes already clustered. Finally, the method is suitable for real world applications where large number of lanes need to be analyzed.

6. Conclusions

This research considers the problem of clustering lanes of demand in freight logistics networks. This is motivated by the economies of scope achieved by important logistics clusters implemented over the world. Demand clustering is relevant for flexible transporters that need to identify groups of synergetic lanes. These lanes should be profitable under uncertain volumes and prices. Empty-trip reduction is critical to achieve this goal because it considerably decreases operational costs. Furthermore, this phenomenon mitigates negative externalities to society. The clustering problem is approached from a truckload (TL) perspective. TL is the most popular and flexible type of operation for freight transportation.

Demand clustering in logistics networks is important for several reasons. First, it facilitates the analysis and prioritization of demand for TL carriers, which is essential to detect new business opportunities that can be included into their current networks efficiently. Thus, clusters have to be carefully built in order to add synergies that reduce empties and increase profits. Furthermore, optimizing routing and scheduling over the complete network covered by large carriers is computationally demanding. An appropriate clustering approach is vital to detect sub-networks that can be optimized efficiently. Finally, knowledge about lanes that perform well when served together is important to develop pricing and revenue management strategies that add value to the business of their clients, i.e., shippers. For example, two lanes from two separate shippers served in isolation would be individually expensive. However, if economies of scope are achieved and they are part of the same cluster, the carrier can price them lower without monetary loses. This makes the current service competitive (low price), and reduces transportation expenses for the shippers.

This paper proposes a novel algorithmic approach to cluster lanes of demand, which is based on dependent sampling over historical data and a series of network transformations. Briefly, Samples for price and volume are collected using the Latin-hypercube technique. A profit maximization linear program is solved to find the optimal distribution of trucks associated to each sample. Based on these flows, trip-chains are mined to determine the bilateral utility of synergetic lanes. Finally, these utilities are used to populate an interconnectivity network, which is explored with a community detection algorithm to cluster demand lanes.

The contributions of this paper to the literature are (1) proposing a novel clustering framework to consider interdependencies between lanes, (2) incorporating market prices in a revenue management fashion, (3) considering the interrelation and variability of lane volumes and prices, (4) developing an algorithmic approach that is computationally efficient.

Numerical experiments show the importance of the method. Geographic nearness is not the only attribute to consider when clustering demand in logistics networks. The contemporaneous bilateral utility determined by the profit of serving lanes in the same trip-chain is an accurate metric of proximity that takes into account the different dimensions of this complex problem. Additionally, this paper shows that lanes present synergies at different levels, i.e., in a hierarchical fashion. Thus, carriers can analyze the opportunities of serving combinations of lanes with different priorities, which is important for decision making in complex networks. Consequently, in some cases, it is better not to consider some lanes that are in the vicinity of others but do not contribute to their local synergy. The model is scalable for real world applications.

Research hereby can be extended in several directions. First, this work demonstrates that community detection algorithms can be used in logistics problem -specifically demand clustering- using an appropriate definition of the network to cluster. Although the model focusses on the TL market, further research can be developed accounting for modes that not only benefit from economies of scope/frequency but also scale/density. This is the case of consolidated operations, e.g., less-than-truckload (LTL). Accounting for such economies requires the development of appropriate methods to determine bilateral utilities so that community detection is applied in the proper network. Second, although the proposed linear program (LP) is sufficient to capture synergies between lanes, the model can be improved adding other operational constraints in Module 2 if this is required. Practically, any possibility can be explored and complexity will change as a function of the complexity of the implemented approach. Numerical results show that this module roughly contributes to 46% of computational time. So improvements can considerably increase the performance of the overall algorithm. Third, there are considerable similarities between the current LP and the well-known minimum-cost flow (MCF) problem. Framing the LP as a MCF problem will improve performance significantly as several efficient solution algorithms exist for it, e.g., network simplex (Ahuja et al., 1993). Forth, algorithmic efficiency can be improved by developing new efficient approaches in Module 3 to find tours and update interconnections. Currently, this module contributes to roughly 53% of overall computational time. The fundamental properties of efficient algorithms that explore cycles in networks can be approached with this purpose. For example, the efficient Tarjan's algorithm (Tarjan, 1972) finds strongly connected components in directed networks. It determines groups of nodes that are reachable from each other base on arc topology. This concept is similar to the trip-chains analyzed in Module 3. Fifth, the proposed model is static. Therefore, time compatibility, i.e., consideration of time windows, is not approached. Future research can approach the dynamic characteristics of this problem. Finally, this initial framework can be extended to include a pricing management module that considers willingness to pay of the shippers and a cost management module that proposes efficient vehicle routing algorithms with different operational constraints, e.g. in-vehicle consolidation.

References

Ahuja, R.K., Magnanti, T.L., Orlin, J.B., 1993. Network Flows: Theory, Algorithms, and Applications. Prentice Hall, New Jersey.

Andraski, J.C., 2010. Managing Customer Expectations. Voluntary Interindustry Commerce Solutions (VICS). Spring 2010 Symposium. Lehigh University. <http://www.lehigh.edu/~inchain/documents/VICS.pdf> (accessed 26.08.14).

Belson, K., 2010. Keeping Trucks Full, Coming and Going. The New York Times, April 22, page F2, New York.

Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. J. Stat. Mech: Theory Exp. 2008 (10), P10008.

Bodin, L., Golden, B., 1981. Classification in vehicle routing and scheduling. Networks 11 (2), 97–108.

Boile, M., Theofanis, S., Özbay, K., 2011. Feasibility of Freight Villages in the NYMTC Region. Freight and Maritime Program, Rutgers, the State University of New Jersey.

Bowerman, R.L., Calamai, P.H., Brent Hall, G., 1994. The spacefilling curve with optimal partitioning heuristic for the vehicle routing problem. Eur. J. Oper. Res. 76 (1), 128–142.

Cao, B., Glover, F., 2010. Creating balanced and connected clusters to improve service delivery routes in logistics planning. J. Syst. Sci. Syst. Eng. 19 (4), 453–480.

Caplice, C., 1996. An optimization based bidding process: a new framework for shipper-carrier relationship. Ph.D. Dissertation. Massachusetts Institute of Technology, Cambridge, MA.

Caplice, C., Sheffi, Y., 2006. Combinatorial auctions for truckload transportation. In: Cramton, P. et al. (Eds.), Combinatorial Auctions. The MIT Press, Boston.

Chang, T.S., 2009. Decision support for truckload carriers in one-shot combinatorial auctions. Transport. Res. Part B: Methodol. 43 (10), 914–933.

Clauset, A., Newman, M.E.J., Moore, C., 2004. Finding community structure in very large networks. Phys. Rev. E 70 (6), 066111.

De Angelis, L., 2011. A fall in average vehicle loads. Eurostat. European Comission. <http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-SF-11-063/EN/KS-SF-11-063-EN.PDF> (accessed 27.08.14).

Dondo, R., Cerdá, J., 2007. A cluster-based optimization approach for the multi-depot heterogeneous fleet vehicle routing problem with time windows. Eur. J. Oper. Res. 176 (3), 1478–1507.

Ergun, Ö., Kuyzu, G., Savelsbergh, M., 2007. Bid price optimization for simultaneous truckload transportation procurement auctions. Technical Report, Georgia Tech.

Fan, P., Haran, J.G., Dillenburg, J., Nelson, P.C., 2006. Traffic model for clustering algorithms in vehicular ad-hoc networks. In: Proceedings of the IEEE Consumer Communications and Networking Conference 2006 (CCNC2006), pp. 168–172.

Fortunato, S., 2010. Community detection in graphs. Phys. Rep. 486 (3), 75–174.

Girvan, M., Newman, M.E.J., 2002. Community structure in social and biological networks. Proc. Natl. Acad. Sci. 99 (12), 7821–7826.

Huang, G.Q., Xu, S.X., 2013. Truthful multi-unit transportation procurement auctions for logistics e-marketplaces. Transport. Res. Part B: Methodol. 47, 127–148.

Iman, R.L., Campbell, J.E., Helton, J.C., 1981. An approach to sensitivity analysis of computer models. I – Introduction, input, variable selection and preliminary variable assessment. J. Qual. Technol. 13, 174–183.

Jara-Diaz, S.R., 1981. Transportation cost functions: a multiproducts approach. Ph.D. Dissertation. Massachusetts Institute of Technology. Cambridge, MA.

Jara-Diaz, S.R., 1983. Freight transportation multioutput analysis. Transp. Res. 17A (6), 429–438.

Kuyzu, G., Akyol, Ç.G., Ergun, Ö., Savelsbergh, M., 2015. Bid price optimization for truckload carriers in simultaneous transportation procurement auctions. Transport. Res. Part B: Methodol. 73, 34–58.
Lee, C.G., Kwon, R.H., Ma, Z., 2007. A carrier's optimal bid generation problem in combinatorial auctions for transportation procurement. Transport. Res. Part E: Logist. Transport. Rev. 43 (2), 173–191.
Ma, Z., Kwon, R.H., Lee, C.G., 2010. A stochastic programming winner determination model for truckload procurement under shipment uncertainty. Transport. Res. Part E: Logist. Transport. Rev. 46, 49–60.
McKay, M.D., Beckman, R.J., Conover, W.J., 1979. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21 (2), 239–245.
Mesa-Arango, R., Ukkusuri, S.V., 2013. Benefits of in-vehicle consolidation in less than truckload freight transportation operations. Transport. Res. Part E: Logist. Transport. Rev. 60, 113–125.
Metropolis, N., Ulam, S., 1949. The monte carlo method. J. Am. Stat. Assoc. 44 (247), 335–341.
Nagle, T.T., Hogan, J.E., Zale, J., 2011. The Strategy and Tactics of Pricing: A Guide to Growing more Profitably, fifth ed. Prentice Hall, Boston.
Nejad, M.M., Mashayekhy, L., Chinnam, R.B., 2012. Effects of traffic network dynamics on hierarchical community-based representations of large road networks. In: Proceedings of 15th International IEEE Conference on Intelligent Transportation Systems (ITSC), 1900–1905.
Newman, M.E.J., 2003. The structure and function of complex networks. SIAM Rev. 45 (2), 167–256.
Özdamar, L., Demir, O., 2012. A hierarchical clustering and routing procedure for large scale disaster relief logistics planning. Transport. Res. Part E: Logist. Transport. Rev. 48 (3), 591–602.
Patil, G.R., Ukkusuri, S.V., 2007. System-optimal stochastic transportation network design. Transport. Res. Record: J. Transport. Res. Board 2029 (1), 80–86.
Qiong, L., Jie, Y., Jinfang, Z., 2011. Application of clustering algorithm in intelligent transportation data analysis. In: Information and Management Engineering. Springer, Berlin, Heidelberg, pp. 467–473.
Remli, N., Rekik, M., 2013. A robust winner determination problem for combinatorial transportation auctions under uncertain shipment volumes. Transport. Res. Part C: Emerg. Technol. 35, 204–217.
Scherrer, A., Blondel, V., 2014. The Louvain method for community detection in large networks. Université catholique de Louvain, Louvain-la-Neuve, Belgium. 2011. <http://perso.uclouvain.be/vincent.blondel/research/louvain.html> (accessed 26.08.14).
Schönberger, J., 2006. Operational Freight Carrier Planning: Basic Concepts, Optimization Models and Advanced Memetic Algorithms. Springer.
Setar, L., 2013a. IBISWorld Industry Report 48412: Long-Distance Freight Trucking in the US. IBISWorld Inc.
Setar, L., 2013b. IBISWorld Industry Report 48411: Local Freight Trucking in the US. IBISWorld Inc.
Sharman, B.W., Roorda, M.J., 2011. Analysis of freight global positioning system data. Transport. Res. Record: J. Transport. Res. Board 2246, 83–91.
Sheffi, Y., 2012. Logistics Clusters: Delivering Value and Driving Growth. MIT Press, Cambridge, MA.
Sheffi, Y., 2013. Logistics-intensive clusters: global competitiveness and regional growth. Chapter 19 in Handbook of Global Logistics. Springer, New York, pp. 463–500.
Simchi-Levi, D., Chen, X., Bramel, J., 2005. The logic of logistics. Theory, Algorithms, and Applications for Logistics and Supply Chain Management. Springer.
Singh, G., Wenning, B.L., Becker, M., Timm-Giel, A., Gorg, C., 2007. Agent-based clustering approach to transport logistics. In: Proceedings of 2007 IEEE International Conference on Service Operations and Logistics, and Informatics, SOLI 2007, pp. 466–471.
Song, J., Regan, A., 2003. Combinatorial auctions for transportation service procurement, the carrier perspective. Transport. Res. Record: J. Transport. Res. Board 1833, 40–46.
Song, J., Regan, A., 2005. Approximation algorithms for the bid construction problem in combinatorial auctions for the procurement of freight transportation contracts. Transport. Res. Part B: Methodol. 39 (10), 914–933.
Stein, M., 1987. Large sample properties of simulations using Latin hypercube sampling. Technometrics 29 (2), 143–151.
Tarjan, R., 1972. Depth-first search and linear graph algorithms. SIAM J. Comput. 1 (2), 146–160.
Triki, C., Oprea, S., Beraldi, P., Crainic, T., 2014. The stochastic bid generation problem in combinatorial transportation auctions. Eur. J. Oper. Res. 236 (3), 991–999.
Ukkusuri, S.V., 2005. Accounting for Uncertainty, Robustness and Online Information in Transportation Networks. PhD thesis. University of Texas, Austin.
USDOT, U.S. Department of Transportation, 2012. National Transportation Statistics. Research and Innovative Technology Administration. Bureau of Transportation Statistics. Washington, DC. <http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_statistics/index.html> (accessed 26.08.14).
VICS, Voluntary Interindustry Commerce Solutions, 2009. Macy's and Schneider National Filling Empty Miles for Sustainability and Savings. <http://macysgreenliving.com/media/_CustomMedia/EmptyMiles_cs_092809.pdf> (accessed 26.08.14).
VICS, Voluntary Interindustry Commerce Solutions, 2014. VICS (Empty Miles). GS1 North America. <https://www.emptymiles.org/> (accessed 26.08.14).
Wang, X., Xia, M., 2005. Combinatorial bid generation problem for transportation service procurement. Transport. Res. Record: J. Transport. Res. Board 1923, 189–198.
Xu, S.X., Huang, G.Q., 2013. Transportation service procurement in periodic sealed double auctions with stochastic demand and supply. Transport. Res. Part B: Methodol. 56, 136–160.
Xu, S.X., Huang, G.Q., 2014. Efficient auctions for distributed transportation procurement. Transport. Res. Part B: Methodol. 65, 47–64.
Yu, W., Ding, W., Liu, K., 2005. The planning, building and developing of logistics parks in China: review of theory and practice. Chin-USA Bus. Rev. 4 (3), 73–78.