

# Novel Indexes to Measure Competitiveness of Container Shipping Companies

## Konteyner Taşımacılık Şirketlerinin Rekabet Düzeyini Ölçen Yeni İndeksler

Ahmet Selçuk Başarıcı<sup>1</sup> , Tanzer Satır<sup>2</sup> 

### ABSTRACT

The purpose of this study is to propose novel and efficient competitiveness indexes to measure the level of competition among container shipping operators based on a specific region. These indexes should require only basic data, which is full container throughput on the basis of terminal/hinterland and ship operator. This study takes advantages of two methods to propose novel indexes as alternatives to *Herfindahl Hirschman Index* (HHI), which is very popular to measure level of competition. Originally named *Competition-based Overall Similarity Measurement Index* (COSMI) and *Entropy Competitiveness Index* (ECI) utilize overall similarity measure from clustering analysis and entropy methodologies, respectively. Both indexes have been proposed with two variants for each. COSMI200+ ignores the throughput of each SO having an amount less than 200 Twenty-foot Equivalent Units (TEUs), but COSMITOP5 takes into account only the top 5 SOs in terms of local throughput in a hinterland. ECI-JOINT includes a joint entropy coefficient which is constant for each hinterland, but ECI-VAR takes into account a variable entropy coefficient defined by the number of ship operators in each hinterland. Analyzing a dataset for the terminals located in Turkey, the Entropy Competitiveness Index (by means of ECI-JOINT variant) has been exhibited as a good alternative to HHI with a great correlation coefficient with it: 0.97. Theoretically, Competition-based Overall Similarity Measurement Index (by means of COSMITOP5 variant) seems a promising method, but it is highly affected by outliers and inconstant numbers of ship operators per route, indicating a moderate correlation coefficient with HHI: 0.45.

**Keywords:** Cluster analysis, Competitiveness index, Entropy, Maritime container transportation, Overall Similarity Measurement

### ÖZ

Bu çalışmanın amacı, belli bir bölge bazlı olarak konteyner gemi operatörleri arasındaki rekabet seviyesini ölçmeye yarayan yeni ve etkin rekabet indeksleri oluşturmaktır. Bu indeksler terminal/bölge ve gemi operatörü bazlı olarak, sadece başlıca verilerden olan dolu konteyner sevkiyat verilerine ihtiyaç duymalıdır. Bu çalışma, rekabet seviyesi ölçümünde popüler olan *Herfindahl Hirschman İndeksine* (HHI) alternatif yeni indeksler oluşturabilmek için iki farklı yöntemden yararlanmaktadır. Özgün olarak *Rekabet Bazlı Toplam Benzerlik Ölçüsü İndeksi* (COSMI) ve *Entropi Rekabet İndeksi* (ECI) olarak adlandırılan bu indeksler, sırasıyla kümeleme analizi toplam benzerlik ölçüsü ve entropi yöntemlerinden yararlanmaktadır. Çalışmada her iki indeksin ikişer uyarlaması incelenmiştir. COSMI200+, bir lokal bölgedeki 200 TEU'dan daha az yükleme-tahliye performansı gösteren gemi operatörlerini gözardı ederken, COSMITOP5 sadece en yüksek performansa sahip 5 gemi operatörünü dikkate almaktadır. ECI-JOINT uyarlamasında sabit olan ortak bir entropi katsayısı kullanılırken, ECI-VAR uyarlamasında her bir lokal bölgedeki gemi operatörü sayısına göre değişen entropi katsayısı kullanılmaktadır. Türkiye'de mukim terminallere ait verilerin analizi, Entropi Rekabet İndeksi'nin (ECI-JOINT uyarlaması ile) HHI ile 0,97 değerinde korelasyon katsayısına sahip olduğunu ve bu indekse iyi bir alternatif olduğunu göstermektedir. Teorik olarak Rekabet Bazlı Toplam Rekabet Ölçüsü İndeksi (COSMITOP5 uyarlaması ile) umut veren bir yöntemi barındırsa da, dışadüşenler ve rota bazlı değişken gemi operatörü sayısı nedeniyle HHI ile 0,45 değerinde orta dereceli bir korelasyon katsayısına sahiptir.

**Anahtar Kelimeler:** Deniz konteyner taşımacılığı, Entropi, Kümeleme analizi, Rekabet indeksi, Toplam Benzerlik Ölçüsü

Submitted: 06.01.2020 • Revision Requested: 18.03.2020 • Last Revision Received: 26.03.202 • Accepted: 14.04.2020

<sup>1</sup> **Corresponding author:** Ahmet Selçuk Başarıcı / Kocaeli University, Maritime Faculty, Maritime Transportation Engineering, Kocaeli, Turkey  
E-mail: [selcuk.basarici@kocaeli.edu.tr](mailto:selcuk.basarici@kocaeli.edu.tr), ORCID: 0000-0002-2320-8694

<sup>2</sup> Tanzer Satır / Istanbul Technical University, Maritime Faculty, Maritime Transportation and Management Engineering, Istanbul, Turkey  
E-mail: [tsatir@itu.edu.tr](mailto:tsatir@itu.edu.tr), ORCID: 0000-0002-5308-56808

**Citation:** Basarici, A. S., Satır, T. (2020). Novel indexes to measure competitiveness of container shipping companies. *Journal of Transportation and Logistics*, 5(1), 29-53. <https://doi.org/10.26650/JTL.2020.0001>

## 1. Introduction

Container transportation provides advanced levels of punctuality, reliability, and security. Today, many types of cargo are transported in containers because of shorter transit times and less transferring costs (Karam & Eltawil, 2016, p.2). The size of the container traffic, approximately, corresponds to one third of global cargo traffic (Rathnayake & Wijeratne, 2012, p.363). At the global level, a large number of container shipping companies, i.e. ship operators (SOs), establish alliances in point of slot sharing on each other's ship, on the one hand, they make tough competition with each other to secure cargo for their containers.

Measuring the level of competition for terminals or hinterlands is important to determine both the local differences occurred in time and simultaneous changes in various regions. These findings are obligatory to obtain for taking measures primarily against cartelization and for making analysis about terminal investments, container operations, and so on. This study aims at proposing a reliable and meaningful index, which is compatible with the goals of the sectoral studies, for competition theory. One of the best ways to measure the level of competition is to focus on market shares or the amounts of throughput handled by each SO on the basis of a specific location. Today, the only well accepted index taking advantage of using basic data such as market shares is the *Herfindahl Hirschman Index* (HHI). It is benefited even by some prominent competition authorities widely (Miller, 1982, p.593); however, it does not consist of any satisfactory logic behind its methodology. Any beneficiary to calculate a competitiveness index by means of a simple set of data should be able to have alternative methods to achieve it. This paper aims at answering the question, "is there any simple, valid and, reliable alternative method to calculate a competitiveness index by means of a basic dataset?" Here, in this paper, two indexes have been proposed. The methods they rely on are known well, but they have never been adapted into competition theory till now. The first of these, which is named as *Competition-based Overall Similarity Measurement Index* (COSMI), examines the SOs by means of clustering analysis, assessing them based on similarities of their throughput in a hinterland. The other proposed competitiveness index adapted from Entropy theory, which is named as *Entropy Competitiveness Index* (ECI), takes the advantage of uncertainty concept.

Similarity concept has been widely used in many fields. Some of them are image identification, finding information (Khazaeli, 2013; Sicre, 2011; Vander Meer, 1997; Diplaros, 2007; Wilson, 2008; Fauquier & Boujema, February 2003), data mapping (BAE Systems, 2007; Lange, 2013; Vendrig, 2002), sorting (Ye et al., 2016), missing data estimation (Cai, 2016), harmony in music recording test (Robertson, 2013), determination of image quality (Galea et al., 2012), and metric index creation (Novak et al., 2012). These studies indicate that the methodology of cluster analysis includes several methods with several combinations. This study struggles to make the most appropriate combination for a competitiveness index to be utilized in container shipping.

The level of uncertainty in point of which SO secures a cargo in a specific region can give an idea on the level of competition in this region. In the literature, uncertainty is measured through the concept of entropy. Entropy originally had been related to physics, thermodynamics, but it was adapted into managerial uncertainty theory later on (Baray,

2003, p.7). Here, in this paper, entropy has been examined to understand if it is feasible to propose as a competitiveness index.

Both cluster analysis and entropy have been proven in different disciplines and fields by several researchers. This paper adapts them into maritime container transportation to test. This study examines them along with HHI by means of a local dataset. The papers relevant to competitiveness index have been discussed in Section 2, 'Literature review'. Section 3, 'Methodology', explains the concepts of cluster analysis and entropy. Section 4 discusses a case study in Turkish terminals, and then, a statistical dataset is analyzed in Section 5, 'Results and discussion'. Finally, the findings of the study, the implications, the limitation of the research, and the future research directions are presented in Section 6, 'Conclusions and recommendations'.

## 2. Literature Review

This section has been split into 4 sub-sections: 'Competitiveness index', 'Similarity in clustering theory', 'Entropy theory', and 'The gaps in the literature about competitiveness index'. The sub-section 'Competitiveness index' views the literature in general. The following two sub-sections examine to transfer the relevant content properly into COSMI and ECI. The sub-section 'The gaps in the literature about competitiveness index' discusses the reason why COSMI and ECI are proposed to analyze the dataset in this study as competitiveness indexes.

### 2.1. Competitiveness Index

Today, it is seen that there are very few options in terms of methodology measuring competition in liner shipping. The United Nations Conference on Trade and Development (UNCTAD) determines and publishes the Liner Shipping Connectivity Index (LSCI) values for the coast states since 2004. LSCI compares the commercial competitiveness of a country in terms of logistics and transportation. This index expresses the strength of connection of a country with the ports located in other countries by means of a few parameters (UNCTAD, 2015, p.39). LSCI does not involve in comparing the level of competition in different ports. Bartholdi et al. (2016) proposed Container Port Connectivity Index (CPCI) as an alternative to LSCI. CPCI uses the Hyperlink-induced Topic Search (HITS) algorithm to make the search engine ranking of websites on the internet. It provides to process both throughput of imports and exports simultaneously. The HITS algorithm inspires to combine both sides of foreign trade in a single system.

The studies released apart from container transportation can give an idea on how to examine the level of competition. Huggins (2003, p.91) aimed to assess relative economic competitiveness, scoring and sequencing the residential areas in the UK within a single index under measurable criteria. It relies on the data consisting of the number of companies per capita, knowledge-based business, activity rates, GDP per capita, full-time salaries, and unemployment rates.

Clark (2004, p.9) revealed a competitiveness index (Equation 1) to perceive the investment climate in transition economies. This index represents the level of competition based on sector, country, and company.

$$CI_{ijk} = \delta_1 T_{jk} + \partial_2 CP_j + \partial_3 IC_{jk} + \beta x_{ijk} + \gamma z_j + \lambda_j + \eta_k + \varepsilon_{ijk} \quad (1)$$

The competition index used in the analysis is an index representing the amount of competition that firm  $i$  in country  $j$ , and sector  $k$  faces. High values of the index refer to a highly competitive environment. The selected variables are the variables that define the investment climate. The main variables are the *customs tariff rates* ( $T$ ) being controlled for trade policy, *the competition policy index* ( $CP$  - European Bank for Reconstruction and Development competition policy index), in which higher values represent fewer barriers to entry and better enforcement of stronger laws, and *the investment climate indexes* ( $IC$ ). Several variables related to the investment climate may affect the level of competition. These variables are related to finance, soft budget constraints, infrastructure indexes, and burden of regulations. In addition to the main variables, dummy variables representing countries ( $\lambda_i$  and sectors ( $\eta_k$ ) have been added into the regression. Some regressions might consist of country control dummy ( $z_j$  instead of country dummy. Moreover, the company-level control variables ( $x_{ijk}$ ) have been also taken into account in the regression. The symbols  $\delta_1$ ,  $\partial_2$ ,  $\partial_3$ ,  $\beta$  and  $\gamma$  represent the coefficients in the regression.

The Netherlands Competition Authority developed a unique competitiveness index particularly for the detection of cartel structures (Petit, 2012). After performing an extensive literature review, this model gathered 9 indicators under 4 categories constituting the competitiveness index: *the organization level* (the number of chambers), *the prices* in the Netherlands and in the European Union, *condensation* (condensation based on market share and symmetry measurement, the number of companies, and import rate), and *dynamics* (market growth, dissolution rate which means young and disappearing players in the market, the survival rate which describes the ratio between the companies who are in the market in last 4 years, the average number of all companies in the market in the same period, and research and development rate). The absolute values of each indicator are converted to numbers between 0 and 1. By this way, a value is transformed to a comparative value. If the value approximates to 1, the competition level is accepted to decrease, and subsequently, it leads to a collusive oligopoly. Petit (2012) shared the results for several sectors.

HHI measures the level of competition in a market by considering the market shares of the competitors. It can be used to analyze the degree of market concentration in a certain sector. The index value is determined by summing the squares of market shares of each firm on a percentage basis. The highest index value is detected in a monopoly market. If a single company dominates the market with a 100% market share, the index value indicates the score 10,000 (Competition Authority – Turkey, 2018). When the level of competition increases, the index value decreases. When the number of competitors increases, the index value, which is inversely proportional, decreases. An increasing index value along with increasing competitiveness requires a transformation by taking the reciprocal of the HHI value or by subtracting it from 1. Since 1982, the United States Department of Justice Antitrust Department benefits from HHI (Miller, 1982, p.593). Because it is simple, effective and proven, leading and reputable institutions exploit it when developing new scientific methods. Kanagala et al. (2004, p.1277) illustrate the HHI method numerically (Equation 2 and 3).

$$P_i = (AR_i/D) * 100 \quad (2)$$

$$HHI = \sum_{i=1}^n P_i^2 \quad (3)$$

The variables are symbolized by  $n$  for the number of participants in the market;  $P_i$  for the market share of the participant  $i$ ;  $D$  for the total system demand;  $AR_i$  for the amount of resources allocated to the participant  $i$ . Skilling & Zeckhauser (2002) proposed a political competition index (PCI) (Equation 4 and 5). It was set forth simply as an extension of the *political concentration index* released by Herfindahl. It is obtained by subtracting the HHI value from 1 (Skilling & Zeckhauser 2002, p.127).

$$\text{Where } \sum \alpha_i = 1 \quad (4)$$

$$PCI = 1 - \sum \alpha_i^2 \quad (5)$$

The symbol  $\alpha_i$  is the proportion of time in office for party (or coalition)  $i$ . Almeida & Fernandes (2011, p.11) examine the effect of total factor productivity of the developing economies in the long run, where the economic activities are gathered in certain areas (Equation 6, 7, 8).

$$Conc_{jrt} = L_{jrt}/L_{rt} \quad (6)$$

Sector diversity ( $Div$ ) has been demonstrated by HHI.

$$Div_{jrt} = \sum_{k \neq j} \left( \frac{L_{krt}}{L_{rt}} \right)^2 \quad (7)$$

The high values of this index indicate a lower sector diversity. Based on this, the degree of competition in a sector is demonstrated by taking the inverse of the HHI value.

$$Comp_{jrt} = 1/\sum_{i \in Z} \left( \frac{L_{ijrt}}{L_{jrt}} \right)^2 \quad (8)$$

The variables are symbolized by  $Z$  for the set of integers;  $i$  for firm;  $r$  for region;  $t$  for time;  $j$  for sector;  $L$  for total employment and  $Conc$  for industry concentration.

The competitiveness indexes, which have been proposed in this paper are respectively based on clustering analysis and entropy. Herein below, the clustering and entropy theories are examined.

## 2.2. Similarity in Clustering Theory

Clustering analysis is one of the interdependent techniques such as factor analysis. Interdependent techniques focus on only the identification of the structure (Hair et al., 2009, p.553). This analysis investigates the similarity of observation values, which have multiple variables. It assorts some observation values, which have similar characteristics within a homogeneous structure, to a cluster and some others to another cluster. It benefits from the similarity theory while performing this operation. In many fields, it is used for taxonomy of similar observation values. Some of these fields are the personality types in psychology, the consumer behaviour in product usage, and the chemical components in terms of performance characteristics (Dillon & Goldstein, 1984).

Clustering analysis utilizes various clustering procedures. They are commonly classified under two categories: hierarchical and non-hierarchical separation procedures. Hair et al. (2009, p.585) provide an overview in this context. Hierarchical separation procedure gathers the closest observation units under a single cluster, then it connects them each other within a rule. The most popular agglomerative algorithms are single-linkage, complete-linkage, average linkage, centroid method, and Ward's method. The non-hierarchical procedures determine the number of clusters from the beginning, and chooses the starting (or seed) point for each cluster. The basic approach is to form a cluster including the closest observation units surrounding a seed point. Hierarchical clustering analysis is quite suitable for almost all kinds of research questions. However, one of the most prominent disadvantages of hierarchical methods is its sensitivity to outliers.

Picard & Franc (2003, p.181) examine the diversity of rainforest trees. They grouped the tree species. The paper proposes a new grouping method. This method compared Ward's method and Euclidean distance which uses hierarchical clustering analysis. Wei et al. (2006) proposes to assort documents on the basis of content. Thereby, the documents can automatically be assorted based on pre-determined categories. In addition to a hybrid document clustering technique, it utilizes similarity measurement improved from a method of vector distance measurement. Similar observations are assorted into the same clusters by the hierarchical clustering approach. Both these studies are good examples to demonstrate how to benefit hierarchical clustering analysis.

Basically, similarity is measured in two different ways: correlational measurement and distance measurement. *Correlational measurement* means that the value of the variable for each observation unit is compared pairwise with the value of the variable for another observation unit, and then a correlation matrix is formed. On the other hand, *distance measurement* takes into account the vector distance between the observation units. Rouhizadeh (2015, p.10-13) released a composing study about the methodology on vector-based similarity measurement. It scrutinises the use of language within disease of autism. The usage frequency of the words refers to a vectoral distance. In a vector space, if the vectors are close to each other, it refers to the assumption that the meaning of the words is semantically similar. The study expresses five different measurement methods, which are herein described briefly, in terms of the vector space similarity. They are the Jaccard similarity coefficient, Cosine similarity score, Relative frequency measure, BLEU score (Bilingual evaluation understudy), and Vector distance measurement. There are various options in determining the distance measurement: the Euclidean distance (Equation 9), Squared Euclidean distance, Manhattan distance (City-block), Chebychev distance, and Mahalanobis distance (Hair et al., 2009, p.575). Euclidean distance refers the distance between two points in a coordinate plane.

$$Euclidean\ distance = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2} \quad (9)$$

The variables are symbolized by  $X$  for variables on axis of abscissas and  $Y$  for variable on axis of ordinates. The Squared Euclidean distance takes advantage of the Centroid and Ward clustering methods.



Distance measurement does not truly describe similarity, but it represents condensation examined in the competition theory because similarity decreases as distance increases. It is possible to convert the distance measurement, i.e. condensation score, into similarity score. As Skilling & Zeckhauser (2002) and Almeida & Fernandes (2011) convert concentration index into competitiveness index, Turney et al. (2010, p.161) proposed two ways (Equation 10, 11) to convert it: inversion of the distance measurement and subtraction.

$$sim(x, y) = 1/dist(x, y) \quad (10)$$

$$sim(x, y) = 1 - dist(x, y) \quad (11)$$

The operators are symbolized by *sim* for similarity, *dist* for distance, and the variables are symbolized by *x* and *y* in the coordinate plane. Niemann et al. (2012, p.171-172) benefited these conversions in the context of the similarity theory. The study is about business process management supported by information technologies. It comprises a comparison of the process models and disclosure of their similarities. Regardless, the meaning of the word and phrase directories,  $L_1$  and  $L_2$ , the similarities are calculated. The distance measurement is accepted as Levenshtein's distance,  $lev(L_1, L_2)$ . The operation (Equation 12) subtracts distance from 1 to reach a similarity measurement.

$$sim^{lev}(L_1, L_2) = 1 - lev(L_1, L_2) \quad (12)$$

On the other hand, Niemann et al. (2012, p.171) also examine the semantic similarity between the words in itself and the sentences in itself. It has been defined as the *word distance* to access the semantic similarity. A word distance metric is based on WordNet (*WN*), which is a lexical database of semantic relations between words. The shortest path between two words,  $w_1$  and  $w_2$  as word-based network, is defined as  $\Delta^{WN}(L_1, L_2)$ . The similarity value (Equation 13) is expressed by the inverse of this distance, i.e. WordNet distance (*WND*).

$$sim^{WND}(L_1, L_2) = 1/\Delta^{WN}(L_1, L_2) \quad (13)$$

The proposal to convert distance measurement into the similarity measurement is extremely important because, in this way, the higher similarity score provides the higher index score.

### 2.3. Entropy Theory

The concept of entropy has been transferred into management science by Shannon (Baray, 2003, p.12). The weighting theory frequently has used entropy method. Some studies related to containerization have utilised it, too. Lee et al. (2012, p.5653) compare the financial statements of four Taiwanese and South Korean SOs. The study takes advantage of the entropy method and the gray relational analysis (GRA). The entropy method was exploited to weight the financial ratios whereas GRA ranked the performance of the SOs. Yang and Shen (2013, p.165) compared conventional and automated container terminals in terms of their operational performances. They implemented the identical methods as it had been released by Lee et al. (2012). Lee et al. (2014, p.246) compared the financial statements of container SOs. In this study, the weighting is implemented by two different ways. One of them is subjectively performed through questionnaires; the other one is

performed objectively, evaluating the financial statements, and then both are compared. Subjective weighting uses a fuzzy logic method: Consistent fuzzy preference relation. The other one benefits from entropy. The study concludes that an objective weighting cannot be performed through a subjective assessment. Su et al. (2016, p.25-27) compared some hub ports in the Far East: Hong Kong, Kaohsiung, and Xiamen, in the context of operational and managerial vulnerability. In the study, the indicators on vulnerability are weighted by the entropy method. The index scores are determined by GRA scores, and finally the ports are ranked by means of Analytic hierarchy process (AHP).

The literature review reveals that the entropy method has not been previously exploited to create an index to measure competitiveness. Entropy is utilized to measure the magnitude of uncertainty. One can discourse that the absence of competition in a market refers to a monopoly market. In such a market, one cannot alledge that it is uncertain. It is certain who will be selected by a customer as the shipping company for transportation. The concept of entropy, which was transferred from physics and information sciences to social sciences, contains characteristical information in the decision matrix (Ömürbek et al., 2016, p.238). Baray (2003) clearly illustrates the logic of the entropy concept.

“Entropy can be explained and understood by the concepts of macro and micro circumstances. Macro circumstance represents a case in which a box is full of molecules which are equally or unequally distributed in both halves of it. On the other hand, micro circumstance represents the number of combinations of each molecule in a macro circumstance. If the molecules are evenly distributed into both halves of a box, the number of combinations in terms of replacement of the molecules becomes the highest. If the identical number of molecules are not evenly distributed into both halves, the number of combination decreases. In other words, if the number of combinations is low, entropy is low (i.e. uncertainty is low), if it is high, entropy is high (Baray, 2003, p.8-9).”

When one adapts this explanation to a competition environment in container shipping, it can be concluded that if there is a single SO in a market, it is certain who secures the cargo, but if there are numerous competing players, it will be relatively uncertain who secures it: it means entropy increases. In other words, the case of high uncertainty refers to increased competition, and on the other hand, the case of less uncertainty refers to a lower level of competition.

The literature about the competitiveness index in shipping contains some gaps to overcome. They are discussed herein below.

#### ***2.4. The Gaps in the Literature About the Competitiveness Index***

The well-known indexes in container shipping LSCI and CPCI are not related to the competitiveness index, but they just focus on the strength of the connection of a country with the rest of the ports in the World in terms of maritime container transportation. HITS algorithm benefited by CPCI seems inadequate in terms of pointing to the rivalry among SOs, but it inspires to combine both sides of foreign trade in a single system.



In the literature, some developed methods such as Huggins (2003) and Clarke (2004) about the competitiveness index keep several indicators in their models, therefore, they seem complicated and require various datasets. Generally, it is not easy and practical to get a wide series of data. The methods proposed by Huggins (2003) and Clarke (2004) apart from shipping aim at disclosing the fact of competitiveness in local production units, and they pursue the local indicators such as unemployment rates and the investment climate. On the other hand, the approach revealed by Petit (2012) seems very attractive as a preferable index in point of the competition among SOs, however, it is relatively complicated and is designed to detect cartelization. The abovementioned methods are doubtlessly precious, but their contents do not fit the aim of this study. Additionally, they require a dozen datasets from various resources, too.

This study struggles to find a simple but substantial way of measuring the level of competition. The practitioners and the academicians should be able to attain an index revealing the level of competition in a region by just having a dataset including the amounts of throughput of each player in a specific terminal, port, etc. Today, the only simple and proven way to measure competition in shipping this way is HHI. On the other hand, proven and well-known methods, which are entropy and a multivariate analysis technique: cluster analysis, have been adapted to this area in this study. The index arising from the cluster analysis is named COSMI, which is inspired by the HITS algorithm combining both sides of foreign trade, imports and exports, in a single system. The other index arising from the entropy method is named ECI.

### 3. Methodology

The section ‘Methodology’ explains the methodologies behind COSMI and ECI respectively. Introductions, the pros and cons of these indexes, their variants, and their implementation steps are elaborated in the following sub-sections.

#### 3.1. COSMI

Overall similarity measure (OSM) is a term of the clustering analysis technique (Hair et al., 2009, p.565). It provides to measure the similarity level of observation values through multiple variables. Each distance between observation values formed by 2 different parameters on a coordinate plane is calculated by means of an appropriate technique, and then it is disclosed as a numerical value. This value is referred as OSM value. Consecutive and multiple calculations generate an OSM value for an observation group. The OSM value of an observation group can be compared with another group. In this way, one can detect which group’s elements are more similar. Clustering analysis seems quite suitable to detect similarities of SOs in terms of their throughput per terminal or hinterland. This paper represents that the two parameters for each observation are the throughput for imports and exports. Hereby, OSM values are calculated based on route, such as Istanbul-Kocaeli hinterland – Far Eastern ports, for both inbound and outbound throughput simultaneously. If an OSM value is relatively small, it means that the similarity, i.e. competition, level in this group is high.

Clustering analysis includes different clustering procedures, different measurement types, and several consecutive methods to find out a valid OSM value. This paper struggles to form the best methods to acquire the most appropriate OSM values. This combination creates COSMI.

“Hierarchical clustering analysis is quite suitable for almost all kinds of research questions. One can benefit the centroid method in hierarchical clustering procedure and can use the Squared Euclidean distance in distance measurement to attain the best results. However, one of the most prominent disadvantages of hierarchical methods is its sensitivity to the outliers. Additionally, the centroid method is sensitive to dissimilar scales and magnitudes (Hair et al., 2009, p.590).”

Therefore, OSM values must be calculated as standardized observation values. Furthermore, the measure of the Squared Euclidean distance can somewhat deal with the outliers and limited number of observation units (Hair et al., 2009, p.575).

The contribution of this paper to the competition literature is to propose an algorithm utilizing the clustering analysis. COSMI struggles to eliminate its weaknesses and adapts it into a competitiveness index. Herein below, the pros and cons of COSMI are discussed.

The advantageous parts of COSMI:

1. Since observation values are created simultaneously, weighting operation based on inbound and outbound throughput is fulfilled inherently.
2. Distant subsets in a coordinate plane reveal the fact that SOs compete within different segments in point of container throughput. When the distance between subsets increases, OSM value increases, therefore, competition level decreases; this is the expected result.
3. Since an OSM value is the sum of the distances between observation points, it is kept constant even if small and large observation values are replaced by axis of symmetry in a coordinate plane. Therefore, OSM is not sensitive to the magnitude of the observation value, and emphasizes the significance of competition even though subsets include a lower amount of throughput.
4. OSM is a well accepted method of the clustering theory.

The disadvantageous parts of COSMI:

1. OSM is sensitive to subsets and outliers even though observation values are standardized.
2. OSM is sensitive to the number of observation units. The more observation units are in a set, the more cumulative distance occurs, accordingly, the OSM value rises. Since the number of SOs differs per route, the OSM value may be dramatically affected.

One has to face a trade off between eliminating the disadvantages of COSMI as much as possible and keeping maximum amount of data in a set. Therefore, two approaches have been adopted: COSMI200+ and COSMITOP5. COSMI200+ ignores the throughput of each SO having an amount less than 200 TEUS for both inbound and outbound traffics

in total. On the other part, COSMITOP5 takes into account only the top 5 SOs in terms of local throughput in a terminal or in a hinterland. COSMI200+ keeps the data in a set as much as possible, whereas COSMITOP5 struggles to eliminate the disadvantages of OSM arising from the additive distance.

Implementation steps of COSMI:

Calculations are made through route-based datasets for each hinterland.

1. a. For COSMI200+: in case the sum of inbound and outbound throughput on the basis of route and SO in a dataset is below 200 TEUS, this part of data is subtracted from the original dataset.  
b. For COSMITOP5: all minor data in a dataset except the data reflecting top 5 SOs per route is subtracted from the original dataset.
2. The dataset per route is standardized (Z score). OSM values are calculated through the centroid method. Distant measurement is Squared Euclidean distance. First, the two closest elements of the cluster and then distance between them are determined. This transaction is repeated for the rest of elements similarly. Distances between clusters can be calculated through the centroids of subsets. The sum of the distances of each subset produces an OSM value per route.
3. Since the result is associated with distance measurement, one has to convert it into the similarity measure by inverting OSM value, obtaining similarity scores on route basis, i.e. COSMI per route. Since the OSM values are over 1, inverting operation is preferred instead of subtracting OSM value from 1.
4. Similarity scores are multiplied by weighting scores of each route and then summed up to obtain COSMI scores on the basis of hinterland. Weighting operation takes into account the throughput of each route.

### 3.2. ECI

The entropy method takes in consideration the market shares of SOs in each route. The logarithm of market share represents the element of surprise (Baray, 2003, p.12). Normalized values are between 0 and 1. The logarithm of a value which approximates 1 is smaller than the logarithm of a value which approximates 0. It means that the more market share a SO has, the less surprise factor emerges to change this market share. In this context, uncertainty per SO is determined by multiplying the market share and the surprise factor for each SO. The sum of the uncertainties per SO is multiplied by the entropy coefficient to calculate the entropy value. Herein below, the pros and cons of ECI are discussed.

The advantageous parts of ECI:

1. The normalization step defines the market shares of each SO. This step, somewhat, simulates HHI.
2. The entropy method does not face an outlier problem because they become insignificant after calculating operations.

The disadvantageous parts of ECI:

1. Entropy cannot transact inbound and outbound data simultaneously as OSM can do. They must be calculated one by one to merge later on.
2. The number of SOs differs in each hinterland, and it may differ in time. Different numbers of SO granting service in each route create a coefficient problem to solve since the entropy coefficient is linked to the number of SO.

The entropy coefficient ensures the entropy value is kept between 0 and 1 (Yang & Shen, 2013, p.163). It depends on the number of SOs per hinterland. While calculating the entropy coefficient, the number of SOs in a matrix must be taken in consideration; however, the number of SOs servicing each hinterland differs. It means that the entropy coefficient of each matrix must be different. On the other hand, each matrix in an index has a relative relationship with one another. This approach causes different numbers of SOs and entropy coefficients per hinterland, emerging a risk of spoiling the entire analysis. A solution is to add all the SOs servicing all hinterlands into the matrices of each hinterland wherein some of them do not have a service. There is another issue supporting this solution: an empirical analysis indicates that when the number of SOs decreases, the value of the entropy coefficient increases, leading to an increase in entropy. However, theoretically, when the number of SOs decreases, one expects that the entropy value, i.e. uncertainty, should decrease; therefore, the entropy coefficient must be kept constant in all matrices on a yearly basis. One should accept that the SOs which do not have a service for some hinterlands are potential competitors for the existing ones. Within the context of the above discussion, the opinion, a *constant entropy coefficient*, has been tested alongside an alternative opinion, a *variable entropy coefficient*. ECI-JOINT accepts a joint entropy coefficient, which is constant for each hinterland. It equals the number of SOs calling at all the hinterlands in the analysis per year. On the other hand, ECI-VAR takes into account a variable entropy coefficient defined by the number of SOs in each hinterland. The contribution of this paper to the literature on ECI is to propose the entropy method to the competition theory, adapting the entropy coefficient.

The implementation steps of the entropy method are detailed below (Yang & Shen, 2013, p.163).

- 1) The decision matrix is normalized (Equation 14).

$$p_{ij} = x_{ij} / \sum_{i=1}^n x_{ij} \quad (14)$$

The variables are symbolized by  $p_{ij}$  for normalized values,  $x_{ij}$  for original matrix values,  $i$  for alternatives,  $j$  for criteria,  $n$  for the number of alternatives.

- 2) The entropy values are calculated for each criterion (Equation 15, 16)

$$e_j = -k \sum_{i=1}^n p_{ij} \ln p_{ij} \quad (15)$$

$$k = (\ln(n))^{-1} \quad (16)$$

The variables are symbolized by  $k$  for entropy coefficient,  $e_j$  for entropy value,  $p_{ij}$  for normalized value.  $p_{ij}$  symbol refers to the market share of a SO in each route. The number of SOs in a market determines the value of the entropy coefficient,  $k$ . The number of alternatives,  $n$ , refers to the number of SOs. In this study, the alternatives are SOs and the criteria are routes.

The entropy values per route for each matrix must be merged to get only one entropy value per hinterland by implementing a weighting operation. Weight of each route respectively for inbound and outbound traffics is determined by dividing the route based throughput by the total throughput of a hinterland. The entropy value is a magnitude between 0 and 1.

Both indexes have been tested by means of a dataset including throughput of container terminals located in Turkey.

#### 4. Case Study for Container Terminals in Turkey

The dataset used in this study is not publicly accessible in Turkey, but has been obtained from the local players. The tables in this study do not include the names of the SOs, intentionally. The local legislation prohibits disclosing corporate data. The dataset is on an annual basis, and it includes inbound and outbound throughput of SOs on route basis. Four-year data covers the period between 2013 and 2016, comprising twenty-foot and forty-foot containers, respectively. It provides to scrutinize Turkey in four hinterlands (Istanbul-Kocaeli hinterland including the terminals: Kumport, Mardas, Marport, Haydarpasa, Evyap port, Yilport, Limas, and Gemlik hinterland including the terminals: Borusan, Gempport, Rodaport, and Izmir-Aliaga hinterland including the terminals: Alsancak, Egegubre, Nemport, and Mersin hinterland including the terminals: Mersin MIP, Assan, Limak). Each hinterland includes 8 routes connecting Turkey to the rest of the world. They are the Far East, Northern Europe, the Mediterranean Sea, Africa, North America, South America, Australia New Zealand, and the Arabian Gulf. Minitab software has been utilized for the calculation of OSM values per route for COSMI.

The paper illustrates how the dataset has been processed for twenty-foot containers in Istanbul-Kocaeli hinterland, in the year 2016 for COSMI200+ (Table 1), COSMITOP5 (Table 2), ECI-VAR, and ECI-JOINT (Table 3, Table 4, Table 5, Table 6, Table 7). The results of these indexes in addition to HHI have then been compared (Table 8).

#### 5. Results and Discussion

Since HHI is well accepted by several authorities worldwide, COSMI and ECI variants have been tested by HHI in terms of structural validation, correlating and corresponding them (Table 1). High value of correlation coefficient means high validity. This study has benefited the Pearson correlation coefficient. HHI correlates with ECI-JOINT highly: 0.97. HHI correlates with ECI-VAR highly, too: 0.86. On the other hand, the results for COSMI are quite far from satisfying the validation. They are - 0.08 for COSMI200+ and 0.45 for COSMITOP5.

COSMI200+ does not correlate with HHI. On the other hand, COSMITOP5 exhibits a better performance. Although the number of SOs competing in some routes is less than

5 in the data set, when almost equalizing the number of SOs in each route per hinterland, a moderate correlation coefficient (0.45) could be attained. Having a similar number of SOs for each hinterland allows to overcome one of the disadvantages of COSMI, which is sensitivity of OSM to the number of observation units. However, COSMI is still sensitive to subsets and outliers. This disadvantage of COSMI has surfaced dramatically. Apparently, it is not the best way to measure the level of competition. Nevertheless, in case of accepting a high degree of data loss, focusing on an equal number of SOs for each route and trivialising small numbers through some arithmetic steps as HHI implements, COSMI can be still promising.

ECI is much more attractive to create an index. As it is discussed in Section 3 ‘Methodology’, different entropy coefficients per hinterland affected the correlation coefficients negatively in terms of validity. As a result, the correlation coefficient for ECI-JOINT is higher, therefore, ECI-JOINT can be evaluated as a good alternative for HHI. It is reliable, and has got a strong background of entropy theory. Besides, ECI-JOINT justifies the validity of HHI, which utilizes a simple method. Both ECI and HHI utilize market shares of the competitors but have entirely different operations in methodology. HHI benefits the power of squaring in maths to highlight the market shares of the dominant players. After the operation, the magnitude of difference between a pair of players gets greater. On the other hand, ECI concentrates on the surprise factor. The market leader secures a cargo with a minimal surprise factor. This operation approximates the possibility of securing cargo of any pair of players whose market shares are different. It means that after the operation, the magnitude of difference between a pair of players gets smaller. However, since the analysis focuses on the correlation of index scores of each matrix per hinterland, the coherence of the results acquired from both methods is great.

The relatively weaker performance of ECI-VAR must be discussed. ECI-VAR stresses that the number of SOs’ services for each hinterland differs on a yearly basis. For example, there have been 25 SOs calling at the Istanbul – Kocaeli hinterland regularly since 2016, but there have been only 16 SOs for the Gemlik hinterland for this period. 25 players competing in a hinterland should provide a more competitive market than a market in which only 16 players compete. This parameter, the variable number of SOs, is expected to tune well the result of a competitiveness index in comparison to keeping the number of SOs constant. The entropy coefficient, which is linked to the number of SOs, could assist to manage it. However, one must keep in mind that the aim of the entropy coefficient is to keep the entropy value between 0 and 1 in a matrix. When it is kept constant, reflecting the maximum number of SOs in all hinterlands, i.e. in several matrices, it works well. If each matrix keeps its own entropy coefficient to interpret the level of competition further, it does not work because theoretical and practical results of the entropy coefficient are opposite while comparing different matrices (See Section 3, Methodology).

The higher scores for both COSMI and ECI reflect higher competition levels. With regard to ECI-JOINT, the Istanbul-Kocaeli hinterland is the most competitive market among all the hinterlands in Turkey in 2013 for both 20-foot and 40-foot full containers. However, the index results indicate that the level of competition in the Istanbul-Kocaeli hinterland



has gradually decreased in time. The Gemlik hinterland exhibits a similar sight. The level of competition for Mersin and Izmir-Aliaga displays an index moving up-and-down. The share of the Istanbul-Kocaeli hinterland throughput in containerized foreign trade of Turkey is 37% for twenty-foot container traffic and 44% for forty-foot container traffic. ECI-JOINT indicates a decreasing trend in the level of competition among SOs. This may arise from the effects of mergers and takeovers in liner shipping during the last decade. Another implication one can conclude is that the higher magnitude of throughput comes true, then the more severe competition is faced in a local region. Istanbul-Kocaeli and Mersin have a considerable part of container traffic in Turkey. Although the level of competition in the Istanbul-Kocaeli hinterland had decreased in time. In 2016, it was still higher than the levels they were in Gemlik and Izmir-Aliaga. Mersin is the toughest hinterland to compete then.

The section ‘Conclusions and recommendations’ addresses the necessity of a reliable competitiveness index, highlights ECI, explains the constraints of this study, and makes recommendations for future studies.

Table 1. COSMI200+ for 20' in the Istanbul-Kocaeli Hinterland, 2016

Ship Operator	Fareast		N. Europe		Intramed		Africa		N. America		S. America		Aus & N. Zea.		Arabian Gulf	
	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound
SO 1	55.548	14.287	30.239	9.871	10.503	10.761	1.640	6.172	1.325	3.349	2.240	3.665	142	3.408	562	4.149
SO 2	23.264	19.266	4.450	1.528	2.763	1.600	535	9.112	668	662	115	2.305	52	912	579	3.372
SO 3	10.581	2.846	6.822	706	895	1.642	157	2.694	631	117	763	745	67	856	973	363
SO 4	8.044	4.962	301	59	112	911	1	665	213	215	13	261	48	374	920	4.728
SO 5	2.963	553	420	73	2.091	4.837	929	2.425	476	557	321	1.475			61	458
SO 6	7.310	2.703	223	305	705	303	5	198	1.301	1.244	0	238				
SO 7	8.042	7.870	2.443	1.663	0	1.604										
SO 8	5.285	2.923	64	239	9.462	16.977										
SO 9	6.893	1.459	1.290	3	161	626										
SO 10	13.186	1.275	462	64	62	362										
SO 11	1.227	1.041			3.458	3.651										
SO 12	1.125	951			2.429	682										
SO 13					376	0										
SO 14					1.207	2.169										
OSM Value per route	10,4527		19,0143		13,2735		4,7730		5,7778		9,2727		7,7368		5,3545	
COSMI per route	0,0957		0,0526		0,0753		0,2095		0,1731		0,1078		0,1293		0,1868	
Weighted COSMI	0,0469		0,0078		0,0146		0,0125		0,0045		0,0032		0,0019		0,0073	
COSMI for Istanbul	0,0987															

Table 2. COSMITOP5 for 20' in the Istanbul-Kocaeli Hinterland, 2016

Ship Operator	Fareast		N. Europe		Intramed		Africa		N. America		S. America		Aus & N. Zea.		Arabian Gulf	
	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound
SO 1	55.548	14.287	30.239	9.871	10.503	10.761	1.640	6.172	1.325	3.349	2.240	3.665	142	3.408	562	4.149
SO 2	23.264	19.266	4.450	1.528	2.763	1.600	535	9.112	668	662	115	2.305	52	912	579	3.372
SO 3	10.581	2.846	6.822	706	3.458	3.651	157	2.694	631	117	763	745	67	856	973	363
SO 4	8.044	4.962	223	305	9.462	16.977	1	665	213	215	13	261	48	374	920	4.728
SO 5	8.042	7.870	2.443	1.663	2.091	4.837	929	2.425	1.301	1.244	321	1.475			61	458
OSM Value per route	5,1486		9,7000		6,0661		4,0555		5,0672		7,8601		7,7368		5,3545	
COSMI per route	0,1942		0,1031		0,1649		0,2466		0,1973		0,1272		0,1293		0,1868	
Weighted COSMI	0,0951		0,0152		0,0319		0,0148		0,0052		0,0037		0,0019		0,0073	
COSMI for Istanbul	<b>0,1751</b>															

Table 3. Throughput per SO, ECI for 20<sup>th</sup> in the Istanbul-Kocaeli Hinterland, 2016

Ship Operator	Fareast		N. Europe		Intramed		Africa		N. America		S. America		Aus & N. Zea.		Arabian Gulf	
	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound
SO 1	55.548	14.287	30.239	9.871	10.503	10.761	1.640	6.172	1.325	3.349	2.240	3.665	142	3.408	562	4.149
SO 2	23.264	19.266	4.450	1.528	2.763	1.600	535	9.112	668	662	115	2.305	52	912	579	3.372
SO 3	10.581	2.846	6.822	706	895	1.642	157	2.694	631	117	763	745	67	856	973	363
SO 4	8.044	4.962	20	12	112	911	0,01	164	213	215	13	261	0,01	0,01	0,01	0,01
SO 5	0,01	0,01	420	73	2.091	4.837	0,01	0,01	476	557	0,01	0,01	0,01	0,01	0,01	0,01
SO 6	7.310	2.703	223	305	705	303	0,01	46	18	0,01	0,01	0,01	1	0,01	1	91
SO 7	8.042	7.870	0,01	0,01	0,01	0,01	1	665	20	9	15	6	5	1	920	4.728
SO 8	0,01	0,01	64	239	9.462	16.977	929	2.425	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
SO 9	6.893	1.459	1.290	3	161	626	5	198	1.301	1.244	321	1.475	48	374	61	458
SO 10	13.186	1.275	462	64	62	362	0,01	0,01	0,01	0,01	0,01	0,01	2	51	2	0,01
SO 11	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	9	47	0,01	0,01	0,01	0,01	0,01	32
SO 12	2.963	553	34	6	0,01	0,01	2	0,01	1	0,01	0,01	0,01	0,01	13	8	5
SO 13	0,01	0,01	2.443	1.663	0,01	0,01	0,01	0,01	0,01	0,01	0,01	238	0,01	0,01	0,01	0,01
SO 14	0,01	0,01	22	0,01	3.458	3.651	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
SO 15	0,01	0,01	17	16	2.429	682	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
SO 16	5.285	2.923	301	59	376	0,01	0,01	38	29	5	0,01	0,01	2	15	0,01	0,01
SO 17	0,01	0,01	0,01	0,01	1.207	2.169	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
SO 18	0,01	0,01	0,01	0,01	0,01	0,01	32	50	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
SO 19	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	5	116	1	0,01
SO 20	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	32	1	0,01
SO 21	1.227	1.041	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
SO 22	1.125	951	0,01	0,01	0,01	0,01	0,01	0,01	0,01	1	0,01	0,01	0,01	19	0,01	0,01
SO 23	0,01	0,01	0,01	0,01	0,01	1.604	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
SO 24	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	16	0,01	0,01	0,01	0,01	0,01	0,01
SO 25	0,01	0,01	0,01	0,01	0,01	0,01	0,01	14	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
Others	2.204	1.126	30	15	268	690	0,01	40	1	0,01	0,01	0,01	3	3	0,01	0,01
<b>TOTAL</b>	<b>145.672</b>	<b>61.262</b>	<b>46.837</b>	<b>14.560</b>	<b>34.492</b>	<b>46.815</b>	<b>3.301</b>	<b>21.618</b>	<b>4.692</b>	<b>6.222</b>	<b>3.467</b>	<b>8.695</b>	<b>327</b>	<b>5.800</b>	<b>3.108</b>	<b>13.198</b>

P.S.: If the throughput value of a SO for a route is zero, this value is accepted as 0,01 so that the calculating operation can be performed properly.

Table 4. Normalised Values, ECI for 20' in the Istanbul-Kocaeli Hinterland, 2016

Ship Operator	Fareast		N. Europe		Intramed		Africa		N. America		S. America		Aus & N. Zea.		Arabian Gulf	
	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound
SO 1	0,387	0,238	0,646	0,679	0,307	0,233	0,497	0,286	0,282	0,538	0,646	0,421	0,438	0,588	0,181	0,314
SO 2	0,162	0,320	0,095	0,105	0,081	0,035	0,162	0,422	0,142	0,106	0,033	0,265	0,160	0,157	0,186	0,255
SO 3	0,074	0,047	0,146	0,049	0,026	0,036	0,048	0,125	0,135	0,019	0,220	0,086	0,207	0,148	0,313	0,028
SO 4	0,056	0,083	0,000	0,001	0,003	0,020	0,000	0,008	0,045	0,035	0,004	0,030	0,000	0,000	0,000	0,000
SO 5	0,000	0,000	0,009	0,005	0,061	0,105	0,000	0,000	0,101	0,090	0,000	0,000	0,000	0,000	0,000	0,000
SO 6	0,051	0,045	0,005	0,021	0,021	0,007	0,000	0,002	0,004	0,000	0,000	0,000	0,003	0,000	0,000	0,007
SO 7	0,056	0,131	0,000	0,000	0,000	0,000	0,000	0,031	0,004	0,001	0,004	0,001	0,015	0,000	0,296	0,358
SO 8	0,000	0,000	0,001	0,016	0,276	0,368	0,281	0,112	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
SO 9	0,048	0,024	0,028	0,000	0,005	0,014	0,002	0,009	0,277	0,200	0,093	0,170	0,148	0,065	0,020	0,035
SO 10	0,092	0,021	0,010	0,004	0,002	0,008	0,000	0,000	0,000	0,000	0,000	0,000	0,006	0,009	0,001	0,000
SO 11	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,002	0,008	0,000	0,000	0,000	0,000	0,000	0,002
SO 12	0,021	0,009	0,001	0,000	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,002	0,003	0,000
SO 13	0,000	0,000	0,052	0,114	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,027	0,000	0,000	0,000	0,000
SO 14	0,000	0,000	0,000	0,000	0,101	0,079	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
SO 15	0,000	0,000	0,000	0,001	0,071	0,015	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
SO 16	0,037	0,049	0,006	0,004	0,011	0,000	0,000	0,002	0,006	0,001	0,000	0,000	0,006	0,003	0,000	0,000
SO 17	0,000	0,000	0,000	0,000	0,035	0,047	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
SO 18	0,000	0,000	0,000	0,000	0,000	0,000	0,010	0,002	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
SO 19	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,015	0,020	0,000	0,000
SO 20	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,006	0,000	0,000
SO 21	0,009	0,017	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
SO 22	0,008	0,016	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,003	0,000	0,000
SO 23	0,000	0,000	0,000	0,000	0,000	0,035	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
SO 24	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,003	0,000	0,000	0,000	0,000	0,000	0,000
SO 25	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
<b>TOTAL</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>

Table 5. ECI-VAR and ECI-JOINT per route for 20' in the Istanbul-Kocaeli Hinterland, 2016

Ship Operator	Fareast		N. Europe		Intramed		Africa		N. America		S. America		Aus & N. Zea.		Arabian Gulf	
	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound
SO 1	-0,367	-0,341	-0,282	-0,263	-0,363	-0,340	-0,348	-0,358	-0,357	-0,333	-0,282	-0,364	-0,362	-0,312	-0,309	-0,364
SO 2	-0,295	-0,365	-0,224	-0,237	-0,203	-0,117	-0,295	-0,364	-0,278	-0,238	-0,113	-0,352	-0,294	-0,291	-0,313	-0,349
SO 3	-0,192	-0,144	-0,281	-0,147	-0,095	-0,119	-0,145	-0,260	-0,270	-0,075	-0,333	-0,211	-0,326	-0,282	-0,364	-0,099
SO 4	-0,162	-0,206	-0,003	-0,006	-0,019	-0,078	0,000	-0,037	-0,140	-0,116	-0,021	-0,105	0,000	0,000	0,000	0,000
SO 5	0,000	0,000	-0,042	-0,027	-0,171	-0,236	0,000	0,000	-0,232	-0,216	0,000	0,000	0,000	0,000	0,000	0,000
SO 6	-0,152	-0,139	-0,025	-0,081	-0,080	-0,033	0,000	-0,013	-0,021	0,000	0,000	0,000	-0,018	0,000	-0,003	-0,034
SO 7	-0,162	-0,266	0,000	0,000	0,000	0,000	-0,002	-0,107	-0,023	-0,009	-0,024	-0,005	-0,064	-0,001	-0,360	-0,368
SO 8	0,000	0,000	-0,009	-0,068	-0,355	-0,368	-0,357	-0,246	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
SO 9	-0,146	-0,090	-0,099	-0,002	-0,025	-0,058	-0,010	-0,043	-0,356	-0,322	-0,220	-0,301	-0,283	-0,177	-0,077	-0,117
SO 10	-0,219	-0,082	-0,046	-0,024	-0,011	-0,038	0,000	0,000	0,000	0,000	0,000	0,000	-0,031	-0,042	-0,005	0,000
SO 11	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,012	-0,037	0,000	0,000	0,000	0,000	0,000	-0,015
SO 12	-0,080	-0,043	-0,005	-0,003	0,000	0,000	-0,004	0,000	-0,002	0,000	0,000	0,000	0,000	-0,014	-0,015	-0,003
SO 13	0,000	0,000	-0,154	-0,248	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,098	0,000	0,000	0,000	0,000
SO 14	0,000	0,000	-0,004	0,000	-0,232	-0,201	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
SO 15	0,000	0,000	-0,003	-0,007	-0,188	-0,062	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
SO 16	-0,122	-0,147	-0,032	-0,022	-0,050	0,000	0,000	-0,011	-0,031	-0,006	0,000	0,000	-0,031	-0,015	0,000	0,000
SO 17	0,000	0,000	0,000	0,000	-0,118	-0,144	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
SO 18	0,000	0,000	0,000	0,000	0,000	0,000	-0,045	-0,014	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
SO 19	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,064	-0,078	-0,003	0,000
SO 20	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,029	-0,003	0,000
SO 21	-0,041	-0,070	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
SO 22	-0,038	-0,066	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,001	0,000	0,000	0,000	-0,019	0,000	0,000
SO 23	0,000	0,000	0,000	0,000	0,000	-0,117	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
SO 24	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,015	0,000	0,000	0,000	0,000	0,000	0,000
SO 25	0,000	0,000	0,000	0,000	0,000	0,000	0,000	-0,005	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
TOTAL	-1,98	-1,96	-1,21	-1,13	-1,91	-1,91	-1,21	-1,46	-1,72	-1,37	-0,99	-1,44	-1,48	-1,26	-1,45	-1,35
ECI-VAR per route	0,614	0,609	0,376	0,352	0,593	0,593	0,375	0,453	0,535	0,426	0,309	0,446	0,459	0,392	0,451	0,419
ECI-JOINT per route	0,581	0,576	0,356	0,334	0,561	0,562	0,355	0,429	0,507	0,403	0,292	0,422	0,435	0,371	0,427	0,396



**Table 6.** ECI-JOINT for 20' in the Istanbul-Kocaeli Hinterland, 2016

	Fareast		N. Europe		Intramed		Africa		N. America		S. America		Aus & N. Zea.		Arabian Gulf		Total
	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	
IST-KOC. 20' Volume	145.672	61.262	46.837	14.560	34.492	46.815	3.301	21.618	4.692	6.222	3.467	8.695	327	5.800	3.108	13.198	420.068
Weight	0,3468	0,1458	0,1115	0,0347	0,0821	0,1114	0,0079	0,0515	0,0112	0,0148	0,0083	0,0207	0,0008	0,0138	0,0074	0,0314	1
ECI per route	0,5807	0,5762	0,3557	0,3335	0,5614	0,5615	0,3547	0,4287	0,5066	0,4027	0,2922	0,4224	0,4346	0,3707	0,4269	0,3962	
Region Score	0,2014	0,0840	0,0397	0,0116	0,0461	0,0626	0,0028	0,0221	0,0057	0,0060	0,0024	0,0087	0,0003	0,0051	0,0032	0,0124	<b>0,5140</b>
ECI-JOINT for Istanbul	<b>0,5140</b>																

**Table 7.** ECI-VAR for 20' in the Istanbul-Kocaeli Hinterland, 2016

	Fareast		N. Europe		Intramed		Africa		N. America		S. America		Aus & N. Zea.		Arabian Gulf		Total
	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	Inbound	Outbound	
IST-KOC. 20' Volume	145.672	61.262	46.837	14.560	34.492	46.815	3.301	21.618	4.692	6.222	3.467	8.695	327	5.800	3.108	13.198	420.068
Weight	0,3468	0,1458	0,1115	0,0347	0,0821	0,1114	0,0079	0,0515	0,0112	0,0148	0,0083	0,0207	0,0008	0,0138	0,0074	0,0314	1
ECI per route	0,6136	0,6088	0,3758	0,3524	0,5932	0,5933	0,3748	0,4530	0,5353	0,4256	0,3088	0,4463	0,4592	0,3917	0,4510	0,4187	
Region Score	0,2128	0,0888	0,0419	0,0122	0,0487	0,0661	0,0029	0,0233	0,0060	0,0063	0,0025	0,0092	0,0004	0,0054	0,0033	0,0132	<b>0,5431</b>
ECI-VAR for Istanbul	<b>0,5431</b>																

**Table 8.** Comparison of competitiveness index results

	Hinterlands	Container	HHI	ECI-VAR	ECI-JOINT	COSMI200+	COSMITOP5
2013	Istanbul-Kocaeli	20'	7.3156	0.6357	.6357	.0867	.1965
	Gemlik	20'	4.8355	0.5952	.5058	.0872	.1356
	Izmir-Aliaga	20'	4.7245	0.5456	.5164	.0780	.1321
	Mersin	20'	6.5132	0.6170	.5980	.1194	.1605
	Istanbul-Kocaeli	40'	6.7830	0.6070	.6010	.0753	.1461
	Gemlik	40'	4.0983	0.5250	.4546	.1205	.1729
	Izmir-Aliaga	40'	5.2368	0.5677	.5373	.1005	.1622
	Mersin	40'	6.8689	0.6267	.6074	.1064	.1580
2014	Istanbul-Kocaeli	20'	5.3922	0.5765	.5534	.0905	.1669
	Gemlik	20'	3.9433	0.5689	.4487	.0904	.1355
	Izmir-Aliaga	20'	4.5385	0.5328	.4931	.1016	.1479
	Mersin	20'	5.5101	0.5468	.5468	.1355	.1964
	Istanbul-Kocaeli	40'	5.1634	0.5529	.5365	.0699	.1388
	Gemlik	40'	3.5081	0.5057	.4173	.1214	.1616
	Izmir-Aliaga	40'	4.2246	0.5256	.4864	.0970	.1572
	Mersin	40'	5.5368	0.5512	.5513	.0812	.1440
2015	Istanbul-Kocaeli	20'	4.7807	0.5662	.5276	.0905	.1647
	Gemlik	20'	3.7601	0.5601	.4338	.0960	.1427
	Izmir-Aliaga	20'	4.3043	0.5144	.4794	.1169	.1709
	Mersin	20'	5.3828	0.5425	.5425	.1030	.1775
	Istanbul-Kocaeli	40'	4.7678	0.5478	.5164	.0759	.1379
	Gemlik	40'	3.2955	0.4795	.3964	.1258	.1538
	Izmir-Aliaga	40'	4.2518	0.5176	.4880	.0957	.1423
	Mersin	40'	5.5866	0.5543	.5494	.1085	.1876
2016	Istanbul-Kocaeli	20'	4.2806	0.5431	.5140	.0987	.1751
	Gemlik	20'	3.1674	0.5107	.4163	.0890	.1214
	Izmir-Aliaga	20'	4.4523	0.5352	.5001	.1150	.1601
	Mersin	20'	5.2584	0.5601	.5601	.0945	.1744
	Istanbul-Kocaeli	40'	4.4973	0.5260	.5039	.1020	.1677
	Gemlik	40'	3.1758	0.4636	.4014	.0868	.1294
	Izmir-Aliaga	40'	4.2342	0.5300	.4953	.1057	.1623
	Mersin	40'	5.7125	0.5739	.5740	.1344	.1936
	<b>Correlation with HHI</b>			<b>0.86</b>	<b>0.97</b>	<b>-0.08</b>	<b>0.45</b>

## 6. Conclusions and Recommendations

Container shipping has been exposed to consolidations in last 20 years, whereas it is still highly competitive. Many SOs disappeared or lost their independence by means of takeovers or mergers, and resulted in the existence of a lower number of independent SOs, especially, in intercontinental strings. Measuring the level of competition between terminals or hinterlands is important to determine both the local differences occurred in time and simultaneous changes in various regions. These findings are obligatory to take measures against cartelization and to make an analysis about terminal investments,

container operations, and so on. This study aims to propose a novel competitiveness index, which is reliable and harbours a meaningful method. Besides, this index must be compatible with the goals of the sectoral studies.

HHI is well accepted by the authorities worldwide to create a competitiveness index because it requires only the market shares of the players. Therefore, it is very simple to practise, but there is not any significant meaning behind it. It just squares the market shares of each player and sums them up to reach a conclusion about the competitiveness level in a market. It is practically useful but a novel index with a substantial background in terms of methodology enriches competitiveness theory. In this context, the clustering analysis, which is a multivariate analysis technique, and entropy have been adapted to create and to test novel indexes. The index based on clustering analysis is named COSMI, and the index based on entropy is named ECI. The magnitudes of inbound and outbound throughput of SOs for a hinterland are enough to create an index by means of these two approaches. Other methods in the literature are complicated and require a series of datasets which are practically quite difficult to collect. Two variants for each index have been tested by HHI, correlating and comparing the results.

In this study, ECI is proposed as a sound alternative to HHI. One of the variants for ECI, ECI-JOINT, correlates with HHI highly. The correlation coefficient is 0.97. It just needs identical datasets as HHI requires. Furthermore, it leans on a strong physics law: the second law of thermodynamics. The concept of entropy has benefited from several disciplines, but it has been adapted into a competition theory for the first time by this study. On the other hand, clustering analysis seems promising to create a competitiveness index at first sight, but COSMI is quite weak due to its sensitivities to the number of observation units and outliers. It requires further improvements to acquire better results; nevertheless, it is certain that a considerable part of dataset must be ignored to overcome its vulnerabilities.

The dataset in this study includes inbound and outbound throughput of SOs at container terminals located in Turkey, and it is based on four hinterlands. One should emphasize the below constraints relating to the dataset. It includes some distant terminals such as Marport/Istanbul and Evyap/Kocaeli in the same hinterland. Infact, benefiting from a data set on the basis of the hinterland instead of on the basis of the terminal is quite logical because neighbouring terminals may serve for the same hinterland, whereas in time, SOs may change their ports of call. Additionally, it must be underlined that the reefer containers and other special types of equipment could not be distinguished and were accepted as standard ones. Nevertheless, this study focuses on the methodologies for a novel index, therefore, these constraints might be ignored.

In liner shipping, especially for container transportation, ECI-JOINT can be considered as a substantial index as an alternative to HHI. For future studies, ECI-JOINT must be tested further alongside HHI by various datasets particularly including more matrices so that it can be well accepted in the literature.

**Peer-review:** Externally peer-reviewed.

**Conflict of Interest:** The authors has no conflict of interest to declare.

**Grant Support:** The authors declared that this study has received no financial support.

**Hakem Değerlendirmesi:** Dış bağımsız.

**Çıkar Çatışması:** Yazarlar çıkar çatışması bildirmemiştir.

**Finansal Destek:** Yazarlar bu çalışma için finansal destek almadığını beyan etmiştir.

## References

- Almeida, R. & Fernandes, A.M. (2013). explaining local manufacturing growth in chile, the advantages of sectoral diversity, *Applied Economics*, 45 (16), pp. 2201-2213.
- BAE Systems (2007). Mediation, alignment, and information services for semantic interoperability (MAISSI): A Trade Study, (Report No. ADA470119), Burlington, Massachusetts: Final Technical Report.
- Baray, A. (2003). Entropi ve karar verme, *Yönetim*, Year: 14, Issue: 44, February 2003, pp.7-21.
- Bartholdi, J.J., Jarumaneeroj, P. & Ramudhin, A. (2016). A new connectivity index for container ports, *Maritime Economics & Logistics*, 18 (3), pp. 231–249.
- Cai, J. (2016). Computational Approaches for Estimating Life Cycle Inventory Data (Master's thesis). Retrieved from <https://deepblue.lib.umich.edu/handle/2027.42/134693?show=full>
- Clarke G. R. G. (2004), How Does the Investment Climate Affect Competition in Transition Economies? papers.ssrn.com. Retrieved March 18, 2017 from SSRN: <https://ssrn.com/abstract=790464> or <http://dx.doi.org/10.2139/ssrn.790464>
- Competition Authority – Turkey (2018). Accessed: 09 January 2018, <http://www.rekabet.gov.tr/tr/Sayfa/Yayinlar/rekabet-terimleri-sozlugu/terimler-listesi?icerik=8bfdbfb0-3cd1-4c6c-82f5-8ae8017267df>
- Dillon, W. R. & Goldstein M. (1984). *Multivariate Analysis Methods and Applications*. New York: John Wiley and Sons.
- Diplaros, A. (2007), Exploiting spatial information for image segmentation and retrieval, (Doctoral dissertation). Retrieved from [https://pure.uva.nl/ws/files/4416614/51736\\_diplaros\\_thesis.pdf](https://pure.uva.nl/ws/files/4416614/51736_diplaros_thesis.pdf)
- Fauqueur, J. & Boujemaa, N. (2003). Region-based image retrieval: fast coarse segmentation and fine color description, *Journal of Visual Languages and Computing* 15 (1), pp.69–95.
- Galea, C., Farrugia, R., Muscat, A. & Zammit S. (2012). Objective quality of experience metrics for television services (Report No. MCA-2012-04). University of Malta: Department of Communications and Computer Engineering.
- Hair, J.F., William, C.B., Babin, B.J., Anderson, R.E., Tatham, R.L. (2009). *Multivariate Data Analysis*. Upper Saddle River, New Jersey: Pearson-Prentice Hall.
- Huggins, R. (2003). Creating a UK Competitiveness Index: Regional and Local Benchmarking, *Regional Studies*, 37:1, 89-96.
- Karam, A. & Eltawil, A.B. (2016). A Lagrangian relaxation approach for the integrated quay crane and internal truck assignment in container terminals, *International Journal of Logistics Systems and Management*, Vol. 24, Issue 1, pp. 113-136.
- Kanagala, A., Sahni, M., Sharma, S., Gou, B. & Yu, J. (2004). A probabilistic approach of Hirschman-Herfindahl Index (HHI) to determine possibility of market power acquisition. *Proceedings of Power Systems Conference and Exposition*, 2004, (pp.1277-1282). The USA: New York, NY, October 10-13.
- Khazaeli, M.A. (2013). Automated Semantic Content Extraction from Images (Doctoral dissertation). Retrieved from [https://digitalcommons.lsu.edu/gradschool\\_dissertations/2697/](https://digitalcommons.lsu.edu/gradschool_dissertations/2697/)
- Lange, D. (2013), Effective and Efficient Similarity Search in Databases (Doctoral dissertation). Retrieved from <https://s3.amazonaws.com/academia.edu.documents/41866560>
- Lee, P.T.W., Lin, C.W. & Chung, Y.S. (2014). Comparison analysis for subjective and objective weights of financial positions of container shipping companies, *Maritime Policy & Management*, 41, 3, pp.241–250.

- Lee, P.T.W., Lin, C.W. & Shin, S.H. (2012). A comparative study on financial positions of shipping companies in Taiwan and Korea using entropy and grey relation analysis, *Expert systems with applications*, Vol.39 (5), pp.5649-5657.
- Lloyd List Maritime Intelligence (2018). Informa, January 16, p.2.
- Miller, R.A. (1982, Fall). The Herfindahl-Hirschman Index As a Market Structure Variable: An Exposition for Antitrust Practitioners, *The Antitrust bulletin*, Retrieved from <http://heinonline.org/HOL>
- Niemann, M., Siebenhaar, M., Schulte, S. & Steinmetz, R. (2012). Comparison and retrieval of process models using related cluster pairs, *Computers in Industry*, 63 (2), p.168-180.
- Novak, D., Batko, M. & Zezula, P. (2012). Large-scale similarity data management with distributed Metric Index, *Information Processing and Management* 48 (5), pp.855-872.
- Ömürbek, N., Karaatlı, M. & Balci, H.F. (2016). Analyzing the performances of automotive companies using entropy based MAUT and SAW methods, *Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler Fakültesi Dergisi*, Vol.31, Issue No.1, pp. 227-255.
- Petit, L. (2012). The Competition Index, the Economic detection instrument of the Netherlands competition authority, papers.ssrn. Retrieved April 12, 2016 from <http://ssrn.com/abstract=1992774>
- Picard, N. & Franc, A. (2003). Are ecological groups of species optimal for forest dynamics modelling, *Ecological Modelling*, 163 (3), pp.175–186.
- Rathnayake, J. & Wijeratne, A.W. (2012). Second container port in Sri Lanka; Hambanthota or Trincomalee: an analysis using the game theory, *International Journal of Logistics Systems and Management*, Vol. 13, Issue 3, pp. 358-378.
- Robertson, H.I. (2013). Testing a new tool for alignment of musical recordings (Master's thesis). Retrieved from <http://digitool.library.mcgill.ca/webclient>
- Rouhizadeh, M. (2015). Computational analysis of language use in autism (Doctoral dissertation). Retrieved from <https://digitalcommons.ohsu.edu/etd/3732/>
- Sicre, R. (2011). Analyse vidéo de comportements humains dans les points de ventes en temps-réel (Doctoral dissertation). Retrieved from <http://www.theses.fr/en/2011BOR14261>
- Skilling, D. & Zeckhauser, R.J. (2002). Political competition and debt trajectories in Japan and the OECD, *Japan and the World Economy*, 14 (2), pp.121-135.
- Su, D.T., Hsieh, C.H. & Tai, H.H. (2016). Container hub-port vulnerability: Hong Kong, Kaohsiung and Xiamen, *Journal of Marine Engineering & Technology*, 15:1, pp.19-30.
- Turney, P.D & Pantel, P. (2010). From Frequency to Meaning: Vector Space Models of Semantics, *Journal of Artificial Intelligence Research* 37, October 2010, pp.141-188.
- Van der Meer, C.A. (1997). A performance analysis of the faugeras color space as a component of color histogram-based image retrieval (Master's thesis). Retrieved from <http://www.dtic.mil/docs/citations/ADA335595>
- Vendrig, J. (2002). Interactive exploration of visual content. The Netherlands: Febodruk BV.
- Wei, C.P., Yang, C.S., Hsiao, H.W. & Cheng, T. H. (2006). Combining preference- and content-based approaches for improving document clustering effectiveness, *Information Processing and Management*, 42 (2), pp.350–372.
- Wilson, W.O. (2008). Immune Inspired Memory Algorithms Applied to Unknown Motif Detection (Doctoral dissertation). Retrieved from <https://pdfs.semanticscholar.org/dd7b/7da0b6ba54be46a24a6ae8a2581984422ba1.pdf>
- Yang, Y.C. & Shen, K.Y. (2013). Comparison of the operating performance of automated and traditional container terminals, *International Journal of Logistics: Research and Applications*, Vol. 16(2), pp.158–173.
- Ye, X. (2016). Automated Software Defect Localization (Doctoral dissertation). Retrieved from [https://etd.ohiolink.edu/pg\\_10?0::NO:10:P10\\_ACCESSION\\_NUM:ohiou1462374079](https://etd.ohiolink.edu/pg_10?0::NO:10:P10_ACCESSION_NUM:ohiou1462374079)