

A data analytic elucidation on the spatial agglomeration of Singapore maritime industry

Lock Yue Chew,^{1,2,*} Ning Ning Chung,³ Wen Xuan Sia,¹ Huynh Hoai Nguyen,⁴ Glenn Sim,^{5,6,7} Alvin Chua,⁷ and Zhongwen Huang⁷

¹*School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore.*

²*Data Science & Artificial Intelligence Research Centre, Nanyang Technological University, Singapore.*

³*Centre for University Core, Singapore University of Social Sciences, Singapore.*

⁴*Institute of High Performance Computing, Agency for Science, Technology and Research (A*STAR), Singapore.*

⁵*School of Global Affairs, King's College London, United Kingdom.*

⁶*School of Global Affairs, Faculty of Social Science and Public Policy, King College London, United Kingdom.*

⁷*Design & Planning Lab, Urban Redevelopment Authority, Singapore.*

(Dated: April 8, 2023)

In this paper, we create three new statistical measures to characterize the spatial patterns and structure of firms and subsectors of industry. The new measures give a direct quantification of the tendency of subsectors to spatially agglomerate, the likelihood of cross-subsector spatial co-agglomerate, and the potential of leading and following behaviour between firms. Unlike conventional measures which deduce possible agglomeration through deviations from the baseline distribution or distance density, our measures study agglomeration in terms of clustering and co-location behaviors to provide a clear physical picture on spatial agglomeration. With each measure quantifying one of the three different ways in which firms agglomerate spatially, their collective use provides a combined perspective into the agglomeration behaviour and location choice of firms that is not available from the conventional measures. We employ this joint application of measures to show that Singapore maritime firms have a propensity to distribute themselves spatially in a manner that allows them to reap the benefits of localization externalities from the agglomeration economy.

I. INTRODUCTION

There is a tendency for maritime industry to agglomerate together [1]. By virtue of being in close proximity to one another, maritime firms may benefit from economies of scales [2]. For example, close spatial proximity reduces the cost for maritime firms to transport goods and provide services to its suppliers and customers. Accordingly, maritime firms located within the same business centres gain access to a larger pool of labour resource that possess the relevant skills

*Electronic address: lockyue@ntu.edu.sg

and knowledge. Productivity improves as maritime firms learn best practices and new knowledge from neighboring firms even as they compete or collaborate with each other [2, 3]. These three Marshallian factors [4] of input-output relationships, pooled labour resources, and information spillover [5–7] are the driving forces that spatially agglomerate firms of the same type or firms of different types within the industry. We term the former *agglomeration* of firms of the same type, while the latter as coagglomeration of firms of different types or simply *co-occurrence*. While there are positive rewards from agglomeration economies, maritime firms may also experience agglomeration diseconomies. Agglomeration diseconomies happen when a clustering of firms leads to higher cost which can result from the need to bear the cost of pollution or the rising cost of infrastructure usage due to congestion [2].

Recently, the presence of agglomeration economies was investigated in the urban setting [1, 8–15]. Studies were undertaken to determine whether economies or diseconomies of agglomeration happen within cities. This relation between economy of scale and spatial agglomeration was in fact captured by urban scaling theory [16]. The effect was revealed by elucidating the manner in which income (wages) scale against the size (density) of population at different geographical scales within a city. The result is both superlinear scaling (economies of agglomeration) and sublinear scaling (diseconomies of agglomeration), and was exhibited, for example, at the local commune level of France [13, 17]. In the context of retail industry, a coagglomeration of restaurants, grocery shops, shopping malls, bars etc has subtle influence on the attractiveness of the locale to the consumer and plays an important part in consumer choice [12].

Specific to this paper is the agglomeration and co-occurrence of firms of the maritime industry in Singapore. Singapore is a city state where top-down urban planning is strategically performed to optimize the efficient use of land which is scarce. Furthermore, Singapore has one of the busiest port in the world, and land planning of its maritime sectors is crucial to the sustenance of this industry. In this respect, the state had made land available for diverse maritime enterprises and has been continually developing and advancing port infrastructures to bring growth to the maritime industry [3]. Fiscal policies such as the Maritime Sector Incentive Scheme by the Maritime and Port Authority of Singapore (MPA) as well as others like the Approved International Shipping Enterprise scheme [18], offers tax exemption and concessionary tax rates to incentivize shipping companies to further expand and enhance their commercial shipping operations in Singapore. This includes extending the ancillary corporate functions and service sectors to create a maritime ecosystem with a broad range of shore-based maritime business ventures. To ensure a supply of skilled professionals, the government has also established several training and education schemes to improve productivity and to drive up innovation. These stimulus programs have the effect of strengthening competitiveness, global connectivity and inclusiveness. They have also led to a comprehensive coverage of maritime activities in Singapore, which has transformed the nation from sole dependence on physical port-based operations to become a vibrant International Maritime Centre. The consequence of these proactive policies by the Singapore government is an integrated maritime industry, which we perceive to be further served endogenously by an agglomeration economy. Research has already shown that agglomeration economies in the maritime industry lowers transportation cost and provide a local environment for positive interactions between firms, which amplifies the process of innovation and the promotion of economic growth in the industry [19]. In fact, benchmark analysis in this respect was conducted in the

maritime industry of Spain, Germany, the Netherlands, and Norway [1]; the result points to the positive impact of agglomeration economies through the sharing of knowledge, risk, and resources.

Thus, from an urban planning perspective, agglomeration can lead to land use efficiency [20]. This is especially important for urban areas which have a high demand of land for diverse developmental needs like Singapore. But not all firm types benefit from such agglomeration effects. The ability to detect which industrial subsectors agglomerate and which do not will help urban planners and policy makers optimize the economic use of land through informed planning. This is true also of firm-type co-location. Co-location of firms is related to spatial co-occurrence of economic activities, which benefits these firms through their implicit collaborative performance of specific urban functions [21]. Moreover, certain industrial sub-types may be the magnet for other sub-types to follow spatially to attain a symbiotic relationship of mutual economic support. The ability to measure and detect such *following behaviour* would help planners spatially distribute firm-types in a self-organized way while optimizing land use efficiency with the associated urban functions.

In order to discern these properties of the agglomeration economies, we need to measure and compare the degree of agglomeration across different industries. For this, empirical data on the economic function and spatial location of firms is necessary for research (or for analysis). Previous analysis of such data had uncovered that low-tech industries tend to agglomerate more than high-tech industries. The comparison was made possible with quantitative measures on agglomeration such as the EG-index [22], the MS-index [23], MTAD [24], and the D -index [25], which are discrete indices because space has been parcelled into discrete regions in their definition. Similarly, spatial co-occurrence between industries has been measured by a index of coagglomeration [26]. There are also the continuous agglomeration indices like the K -function [27] and L -function [28], where spatial description is based on absolute distance measurement. It was observed that the outcome of analysis from these indices on economic agglomeration can be inconsistent. For instance, the EG-index indicates too high an agglomeration of firms in the United Kingdom relative to the K -function. Another crucial element of these indices is that they are derived either from an economic location choice model or from selected statistical tests.

In this paper, we shall define three new quantitative measures for the purpose of uncovering the basic structure of agglomeration economies from empirical data (see Sec. II). Our measures are clustering agglomeration, spatial co-occurrence, and following index. These three measures are computed based on the spatial location of each individual firms. They serve the purpose of elucidating the spatial agglomeration of firms from three different perspectives. Compared to the conventional measures, our measures give a more direct quantification of the spatial agglomeration of firms. Clustering agglomeration determines how firms of the same type cluster, while spatial co-occurrence describes the co-location of firms of different types. Following index illuminates the tendency of firms of one type locating near to the firms of another type. Putting these perspectives together would then provide a holistic view on the spatial morphology and location choice of collections of firms in the industry. In addition, we have adapted these measures for the purpose of big data analysis, and their simplicity has enabled us to perform a fast and efficient computation of large data-sets of firm's locations from diverse industries and commercial entities. They are general enough to be

applicable to any spatial morphological data of cities in the world. We will illustrate the application of our first two measures against the conventional measures of EG-index and K -index on the spatial agglomeration of Singapore's maritime industry to illustrate the differences in their performance (see Sec. III.A and III.B). By using the combined perspectives from our three measures, we draw out empirical details on how a set of Singapore maritime firms would tend to distribute themselves spatially to reap the benefits of localization externalities from the agglomeration economy (see Sec. IV).

II. DATA AND METHODOLOGY

A. Data

The geospatial data analyzed in this study consists of the location of 1,177 firms from the Singapore's maritime industry. The source of the data is from the Singapore maritime services guide [29], which is a one-stop directory of maritime companies offering their services to the Singapore maritime industry. According to the 2017/2018 Singapore maritime services guide, the industry has been classified into 40 subsectors. We have categorized these subsectors (see Table I) into 5 sectors of the Singapore maritime industry to differentiate them. The 5 sectors are: Shipping (S), Port (P), Maritime Services (MS), Commercial Services (CS), and Offshore and Marine Engineering (OME). In our analysis, we have excluded six of the subsectors because these subsectors have less than three firms each.

Index	Subsector	Counts	Index	Subsector	Counts
0	Accounting services (CS)	5	20	Loss adjusting (CS)	7
1	Arbitration centre (CS)	1	21	Marine and offshore engineering (OME)	21
2	Bunker suppliers (P)	60	22	Marine consultancy (MS)	3
3	Bunker Surveyors (MS)	60	23	Marine equipment and supplies (MS)	3
4	Claims Agency (SCMA members) (CS)	1	24	Marine security (MS)	8
5	Classification societies (MS)	12	25	Media and public relations (CS)	6
6	Coal trading (CS)	1	26	Media services (CS)	3
7	Communication Services (CS)	9	27	Port operators and representatives (P)	6
8	Cruise operators (S)	7	28	Recruitment agencies (CS)	9
9	Cruise terminal (P)	2	29	Regional ferry operators (P)	7
10	Education and training (CS)	59	30	Research and consultancy (CS)	35
11	Ferry terminals (P)	4	31	Salvage (P)	11
12	Finance (CS)	31	32	Ship agencies (S)	155
13	Freight derivatives clearing (CS)	2	33	Ship brokers (S)	60
14	Industry organisations and associations (CS)	24	34	Ship builder and repairers (OME)	101
15	Inspection services (MS)	22	35	Ship management (S)	80
16	Insurance (CS)	68	36	Ship owners and operators (S)	173
17	Law firms (CS)	48	37	Ship surveying (MS)	26
18	Law firms (SCMA members) (CS)	29	38	Shipowners - bunker tankers (P)	3
19	Legal services (CS)	1	39	Shipping logistics (S)	14

TABLE I: Subsectors of maritime industry used in the study. The acronym indicates the sector in which the subsector belong to.

B. Clustering Agglomeration

Clustering agglomeration measures the extent firms of the same type cluster. “Type” here refers to the category of the firm in the industry which could encompass the more general industry group to the more specific industry sector. Because different level of type information is encoded in the Standard Industrial Classification (SIC) of the firms, the type of the firm can be efficiently retrieved from data if such SIC information is available. For our analysis on Singapore’s maritime industry, type refers to the subsector of the industry (see Table I) which we obtained from the the Singapore maritime service guide instead of SIC.

To calculate clustering agglomeration, we first perform a clustering analysis using the algorithm of Density-Based Spatial Clustering of Application with Noise (DBSCAN) [30] for a fixed percolation distance d , with d being the Euclidean distance [31, 32]. The fraction of data points f_q within each identified cluster q is subsequently determined. We define a clustering agglomeration index $\alpha(d)$ at distance d through the sum of squares of the fraction f_q as follows:

$$\alpha(d) = \sum_q f_q^2(d), \quad (1)$$

where $0 \leq \alpha \leq 1$. Note that

$$\sum_q f_q = 1. \quad (2)$$

The process is repeated for a range of percolation distance intuited from geospatial and socioeconomic considerations to obtain an average agglomeration $\langle \alpha \rangle$.

Unlike the conventional EG-index [22, 33], clustering agglomeration is studied in a continuous space because administrative boundaries are not required in the clustering analysis. Thus, several issues that accompany the discrete-space assumption are not of concern here [27]. In particular, distance is perceived equally in every part of the continuous space. In addition, our agglomeration index takes the network structure of industry clusters into account. Industry clusters are complex patterns which are not decipherable merely from the distribution of pairwise distance as two firms which are far apart could belong to the same industry cluster and linked through firms between them.

In addition, an arbitrary classification of spatial agglomeration into classes of weak, intermediate, or strong localization is not necessary for the agglomeration index $\alpha(d)$. Instead, one can easily associate the possible clustering pattern(s) to a given value of the agglomeration metric. Let us illustrate what we mean through a few examples. In Fig. 1(a), we have a large cluster that contains 90% of the data points (i.e. $f_1 = 0.9$), while the other five clusters only contain 2% of the data points each. Clustering agglomeration is thus approximately 0.81. On the other hand, a sector with a largest cluster size of $f_1 = 0.7$ (as shown in Fig. 1(b)) is found to possess a similar clustering agglomeration of about 0.5 with a sector that has two equal size cluster: $f_1 = f_2 = 0.5$ (as shown in Fig. 1(c)). Finally, in Fig. 1(d), we show that clustering agglomeration is close to zero for data points which are sparsely distributed.

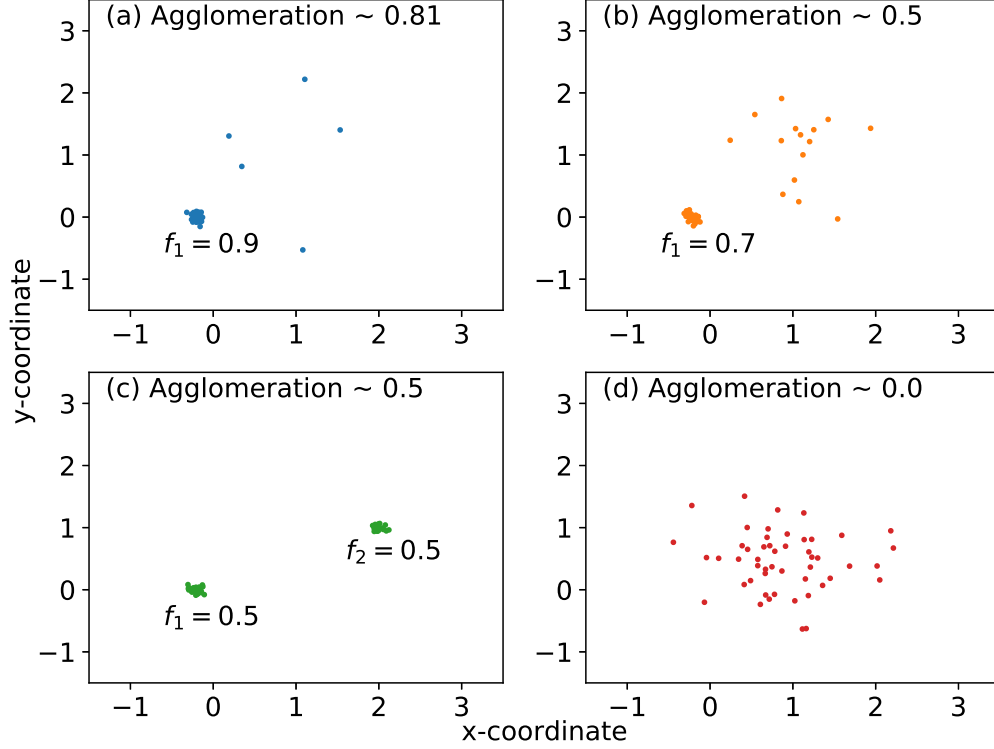


FIG. 1: Conceptual illustration of clustering agglomeration. Note that f_q denotes the fraction of data points in the q -th largest cluster. Here, clustering analysis is done with the algorithm of Density-Based Spatial Clustering of Application with Noise (DBSCAN) for a percolation distance $d < 0.1$.

C. Spatial Co-occurrence

Here, we shall define a measure that captures the tendency of co-location between firms of two different types. The measure will be based on pairwise spatial co-occurrence between spatially distributed data points. First, we form a two-dimensional uniform grid to partition the data points as shown in Fig. 2. The grid size shall range from the micro- to macroscopic scale of the characteristic distances among the spatial points between firms [31]. We then obtain the probability of data points of type i , say, within each cell of the grid. The set of probabilities distributed in the two-dimensional space is then flattened to yield a one-dimensional vector P_i . The spatial co-occurrence between firms of type i and j for a particular grid size is then quantified by the correlation between P_i and P_j as follows:

$$C_{ij} = \frac{\sum_{r=1}^R (P_{ir} - 1/R)(P_{jr} - 1/R)}{\sqrt{\sum_{r=1}^R (P_{ir} - 1/R)^2} \sqrt{\sum_{r=1}^R (P_{jr} - 1/R)^2}}, \quad (3)$$

where $-1 \leq C_{ij} \leq 1$.

When $P_i = P_j$, subsector i co-locates perfectly with subsector j and we have $C_{ij} = 1$. In this case, the probability of finding an i -type and j -type firm in each region r is exactly the same. On the other hand, if the presence of an i -type firm implies the absence of a j -type firm and vice versa, with the difference being uniform across all regions

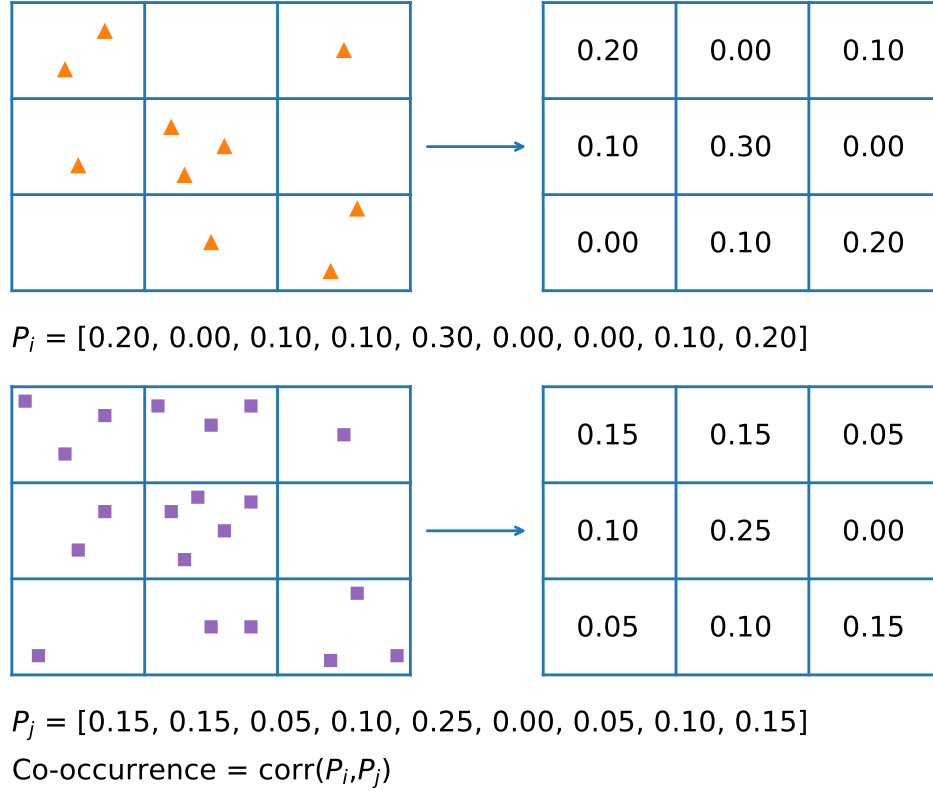


FIG. 2: Conceptual illustration of spatial co-occurrence. Here, the number shown in each cell in the grid on the right gives the fraction of firms found in the corresponding cell in the grid on the left. P_i and P_j are the flattened vectors of the two-dimensional frequencies. $\text{corr}(P_i, P_j)$ denotes the correlation between P_i and P_j .

r , we have $C_{ij} = -1$. An example of this case is when $P_i = [0.5, 0, 0, 0.5]$ and $P_j = [0, 0.5, 0.5, 0]$. Thus, a positive C_{ij} implies that if an i -type firm is present in a certain region, there is a positive chance that a j -type firm is also present in the same region. On the other hand, if C_{ij} is negative, the presence of an i -type firm is likely to indicate the absence of a j -type firm in the same spatial region. When the presence of an i -type firm is neutral to the presence of a j -type firm in the specified region, we have $C_{ij} = 0$. In other words, when C_{ij} vanishes, the probability of finding a j -type firm in a region is totally independent of the probability of finding an i -type firm in that region.

To eliminate the dependence on discrete-space assumption and on grid size, we calculate the correlation with uniform grid at different cell sizes and then take the average to get the spatial co-occurrence $\langle C_{ij} \rangle$. Note however that the cell size of the uniform grid is to be bounded such that the total number of cells in our case is not less than 12. Otherwise, the p -value for a correlation of 0.5 to be significant will be larger than 0.05.

Recently, the co-location of amenities in an urban setting has been studied in a similar manner as our approach [34]. The study, however, involves a neighborhood identification process before the application of Spearman correlation between the number of type- i and type- j amenity to quantify their co-location. Unlike the approach in [34], we choose to study the correlation in a uniform grid to remove any dependence on the underlying firm's distribution. Indeed, the boundaries of the neighborhoods are dependent on the set of data being used in the study. In practice, it is very

hard to obtain a complete dataset that have all the relevant subsectors in an industry. It is also very difficult to justify which subsector to be included or excluded as many industries have a complex relationship with each other. In addition, neighborhoods that are uncovered through the clustering method can appear in diverse sizes. This raises the issue which all measures with administrative or functional boundaries encounter: distance is perceived unequally in different parts of the continuous space of interest.

D. Following Index

The purpose of the following index is to quantify the potential pairwise following behaviour between firms of different types. The idea is that if subsector i is functionally dependent on subsector j and always choose to follow j , one can always find a j -type firm near to an i -type firm. We define following index with i following j as the relative frequency that j is within a distance d of i , i.e.,

$$F_{i \rightarrow j}(d) = \frac{n_{ij}(d)}{N_i}, \quad (4)$$

where N_i denotes the total number of i -type firm and n_{ij} the number of i -type firm near to (within a distance d) at least one j -type firm. As illustrated in Fig. 3, we first examine all i -type firms (shown as triangles) and look at the area within a distance d from each i -type firm to see whether a j -type firm (shown as squares) is present in the circled area. The number of i -type firms with j -type firms in its circle gives $n_{ij}(d)$. Note that in general, following index is asymmetrical:

$$F_{i \rightarrow j}(d) \neq F_{j \rightarrow i}(d). \quad (5)$$

Also, $0 \leq F_{i \rightarrow j}(d) \leq 1$.

It is easy to surmise that the likelihood for a subsector j to be followed by other subsectors increases as the coverage of subsector j in space increases. Consider the extreme case where subsector j is everywhere, we expect $F_{i \rightarrow j} \approx 1$ since there is always a j -type firm nearby no matter which location an i -type firm chooses. We thus propose to look also at the coverage of other subsectors when the following index of a subsector is being examined. To obtain the coverage of a subsector, we partition the two-dimensional space where the firms' reside by a uniform grid as we did in Fig. 2. The grid resolution is based on the characteristic distances between spatial points of firms as discussed in Sec. II C. The coverage of a subsector i is then defined as the number of grid cells with type- i firm over the total number of non-empty grid cells (i.e. grid cells with at least one firm of any type). Based on this definition, we observe that the coverage of subsector i is $3/4$, while that of subsector j is 1 in Fig. 2.

Thus, a high value of the following index $F_{i \rightarrow j}$ is significant only when the coverage of subsector j is also not high at the same time. In such a case, it is very likely that subsector i indeed follows subsector j . In addition, we deem that following behaviour is not significant when the spatial data of i are all clustered at one location and the coverage of i

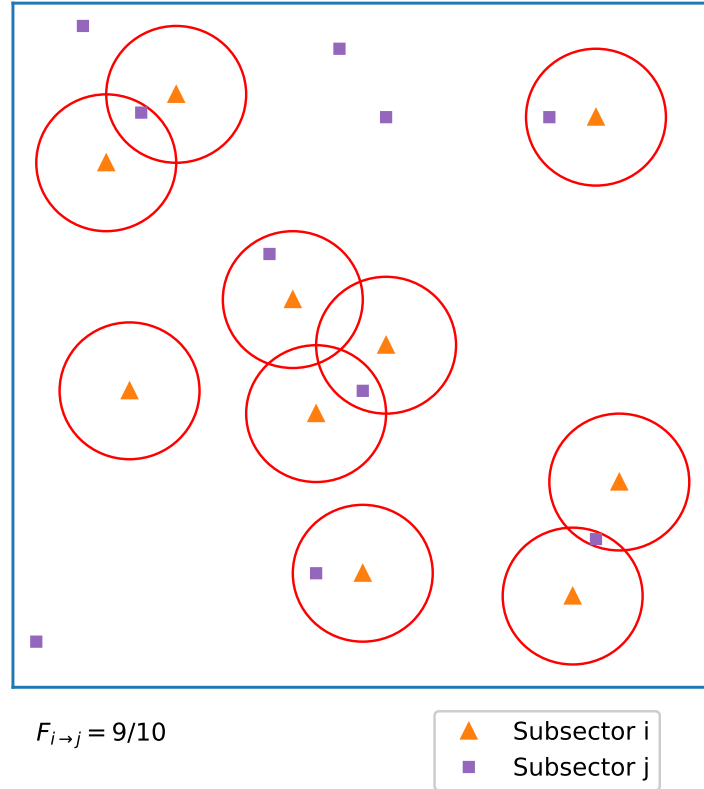


FIG. 3: Conceptual illustration of the following index. We show i -type firms as triangles while j -type firms as squares. Each i -type firm is associated with a circle to illustrate the area that is within a distance d from the firm.

is small. Without the following behaviour occurring consistently throughout space, we cannot conclude that i indeed follows j even when $F_{i \rightarrow j}$ is high. Finally, it is difficult to ascertain following behaviour when data size is small.

It is noteworthy that the colocation quotient [35] provides an alternative way to measure the degree at which one sector is spatially dependent on another sector. Unlike our following index, the dependency is examined through a firm's nearest neighbor – i.e., an i -type firm is dependent on j -type firms if the i -type firm's nearest neighbor is a j -type firm. The measure is based on the belief that the closest firms generally have the greatest influence.

III. RESULTS

A. Clustering Agglomeration

In Fig. 4, we show the average clustering agglomeration $\langle \alpha \rangle$ for 34 subsectors in Singapore's maritime industry at two ranges of length scales: 200 m to 2000 m and 200 m to 5000 m. For each range of length scale, we study the clustering pattern of each subsector at an interval of 200 m to obtain clustering agglomeration at different percolation distances, i.e. $\alpha(d)$. We then perform the averaging operation over the respective range of length scales to obtain

$\langle\alpha\rangle_1$ and $\langle\alpha\rangle_2$, where $\langle\alpha\rangle_1$ is for the range 200 m to 2000 m and $\langle\alpha\rangle_2$ for the range 200 m to 5000 m. These averaging operations reduce the dependence of $\langle\alpha\rangle$ on spatial resolution while uncovering the effect of agglomeration at diverse length scales. A comparison of $\langle\alpha\rangle_1$ and $\langle\alpha\rangle_2$ reveals three classes of industry clustering: (i) subsectors with firms that agglomerate at small length scale, (ii) subsectors with firms that agglomerate at large length scale, and (iii) subsectors with firms that disperse.

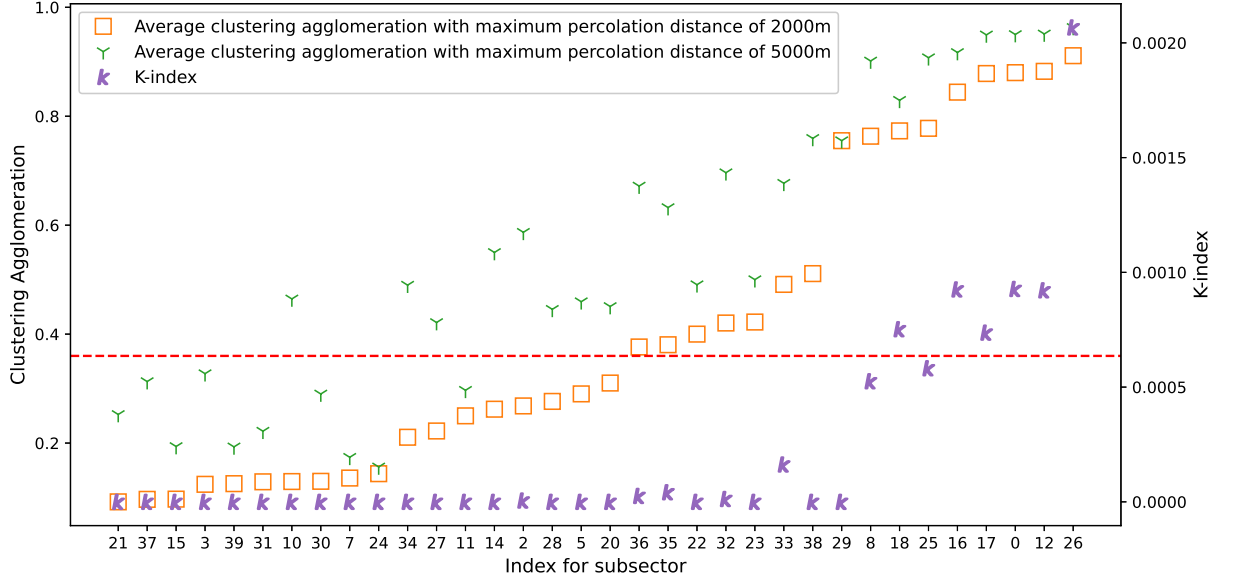


FIG. 4: Average agglomeration for each of the 34 subsectors in Singapore maritime industry under study. Here, two average agglomerations are shown: one averaged over percolation distance from 200 m to 2000 m, while the other averaged over percolation distance from 200 m to 5000 m. The red dotted line is the threshold $\langle\alpha\rangle_T = 0.36$ which segregates between the regime of agglomeration and no-agglomeration.

Subsectors of class (i) have firms that are located at a small distance from each other and are clustered within a small area (see Fig. 5(a)). These firms are always deemed to belong to the same cluster when different percolation distances are used in the clustering analysis. Clustering agglomeration ($\alpha(d)$) calculated for such subsectors are thus high at all percolation distances (d). Some examples of class (i) subsectors are subsectors 26, 12, 0, 17, 16, 25, 18, 8, and 29 (see Fig. 4), which maintain a high value of average clustering agglomeration in both ranges of length scales. On the other hand, subsectors with firms that agglomerate at a larger length scale (i.e., class (ii)) will show a higher average clustering agglomeration value only when the larger percolation distances are considered. In this case, a large deviation is observed between $\langle\alpha\rangle_1$ and $\langle\alpha\rangle_2$. A subsector which belongs to class (ii) is shown in Fig. 5(b) where most of the firms are found to cluster in the west of Singapore. Firms of such subsectors are observed to agglomerate at an intermediate proximity distance with each other. The need for larger space occupancy may be the reason why these firms agglomerate at a larger length scale. In contrast, the average clustering agglomeration $\langle\alpha\rangle_1$ and $\langle\alpha\rangle_2$ are both low for subsectors that belong to class (iii) where firms are observed to disperse in space at both ranges of length scale. One such class (iii) subsector is illustrated in Fig. 5(c).

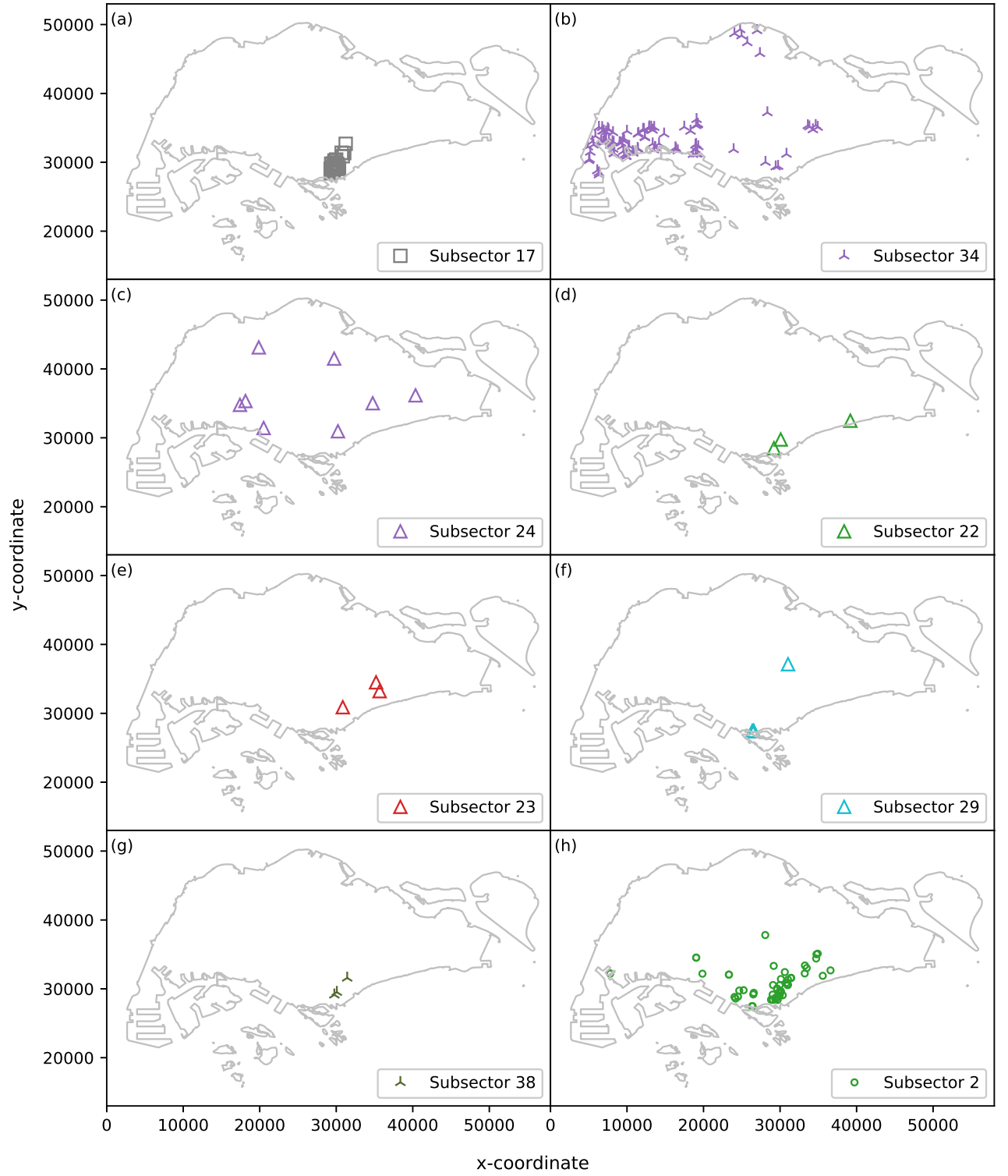


FIG. 5: The spatial distribution of firms in subsectors which: (a) agglomerate at small length scale, (b) agglomerate at large length scale, (c) do not agglomerate, (d)-(g) agglomerate moderately, and (h) distribute in a multi-scale regular pattern.

Next, we compare our calculated clustering agglomerations of the maritime subsectors to a localization measure known as the K -index [27]. The K -index probes the localization of industries through the K -density which gives the estimates of the density of bilateral distances between firms. The K -density employs a Gaussian kernel function f to smooth the density estimate of pairwise distance distribution among N firms at distance d as follows:

$$\hat{K}(d) = \frac{1}{N(N-1)h} \sum_{i_1=1}^{N-1} \sum_{i_2=i_1+1}^N f\left(\frac{d-d_{i_1 i_2}}{h}\right). \quad (6)$$

Note that $d_{i_1 i_2}$ is the distance between firms i_1 and i_2 , and h is the bandwidth of the Gaussian kernel function. For the evaluation of the K -index, we need to compare this K -density of a subsector to a corresponding version of K -density obtained by a random shuffle of all firm locations in the industry. The latter counterfactual random K -density is obtained by means of Monte-Carlo simulation, where we yield its global confidence band. We then determine the upper edge of the global confidence band $\bar{K}(d)$ such that 95% of the random K -density lies below it. This allows us to define a localization index which captures the statistical significance of departures from randomness in the following manner:

$$\kappa(d) \equiv \max(\hat{K}(d) - \bar{K}(d), 0), \quad (7)$$

where $\kappa(d)$ is to be examined from $d = 200$ m to $d = 2000$ m. By taking the largest $\kappa(d)$, we arrive at the K -index, κ .

The K -index κ signifies agglomeration when $\kappa > 0$, which could imply weak to strong agglomeration [27]; and no agglomeration when $\kappa = 0$. In order to compare our clustering agglomeration $\langle \alpha \rangle$ to κ , we need to determine a threshold $\langle \alpha \rangle_T$ such that when $\langle \alpha \rangle > \langle \alpha \rangle_T$, there is agglomeration; and when $\langle \alpha \rangle \leq \langle \alpha \rangle_T$, there is no agglomeration. We found $\langle \alpha \rangle_T = 0.36$ to be a reasonable threshold with the outcome of subsectors' agglomeration comparable between $\langle \alpha \rangle$ and κ . An example of a subsector with $\alpha = 0.36$ is one with three clusters of relative frequencies $f_1 = f_2 = 0.4$ and $f_3 = 0.2$. By setting the clustering agglomeration threshold at 0.36, we thus avoid the situation of classifying a subsector with three dispersed firms (i.e. $\langle \alpha \rangle = 0.33$) as agglomerated.

It is interesting that with $\langle \alpha \rangle_T = 0.36$, 85% of the classifications agree between $\langle \alpha \rangle$ and κ . Disagreements in classification are found only in five subsectors: 2, 22, 23, 29, and 38. Specifically, subsectors 22, 23, 29 and 38 are classified as agglomerated by $\langle \alpha \rangle$ but not by κ . On the other hand, subsector 2 is classified as agglomerated by κ but not by $\langle \alpha \rangle$. Firms in subsectors 22, 23, 29 and 38 are observed to agglomerate moderately in the south or south east of Singapore (see Figs. 5 (d)-(g)). The firms are separated by distances small enough for them to be sufficiently clustered to exceed the threshold $\langle \alpha \rangle_T$ in the percolation analysis. In contrast, the distances between firms are deemed to be 'large' compared to the overall bilateral distance density of the maritime industry as the industry has an overall tendency to concentrate in the central business district. In particular, when a small number of firms are sampled randomly from the whole industry to serve as a baseline for comparison, there is a high chance that the density peak at very small distance. Hence, subsectors 22, 23, 29 and 38 are classified as non-agglomerated compared to the baseline.

On the other hand, firms in subsector 2 are observed to adopt a multiple scale regular pattern [31] (see Fig. 5 (h)). The firms here are found to form clusters that space equally from each other at a length scale of ~ 900 m. Within each cluster, the firms further divide into smaller clusters which are equal distant from each other at ~ 400 m. The overall result is a peak at ~ 900 m in the bilateral distance density of subsector 2, giving rise to a nonzero κ value. Such clusters are nonetheless not well-captured in $\langle\alpha\rangle_1$. It manifests itself however in $\langle\alpha\rangle_2$, as spatial agglomeration becomes apparent at a larger scale.

We have also compared the two measures $\langle\alpha\rangle$ and κ in terms of ranking of the subsectors. Across the 34 subsectors, the Spearman rank correlation between κ and $\langle\alpha\rangle$ is statistically significant at 0.67. We thus conclude that outcome from the two measures agree to a large extent. Note that the Spearman rank correlation between κ and the EG-index is reported to be much smaller at 0.41 for the U.K. industries [27].

In addition to the K -index, we have also computed the D -index for the 34 subsectors in Singapore's maritime industry. Similar to the EG-index, D -index [25] employs a discrete space measure but examines agglomeration using the log-likelihood ratio:

$$D = \sum_r p_{ir} \ln \left(\frac{p_{ir}}{p_{0r}} \right). \quad (8)$$

Note that p_{ir} denotes the probability for a randomly sampled firm of subsector i to locate in region r , while p_{0r} gives the probability for a randomly selected firm (from any subsector) to be in region r . Classification of agglomeration does not match well between the indices when $\langle\alpha\rangle > 0.36$ and $D > 2.6$ are being used to characterize the regime of agglomeration. Furthermore, the Spearman rank correlation between the D -index and $\langle\alpha\rangle$ is -0.21 . Nonetheless, the small negative correlation could happen by chance as it has a p -value of 0.24. Among the disagreement, subsector 24 is of particular interest as it has a relatively high value of D -index (3.45) despite its firms are dispersed over the island, and being dispersed means that this subsector has a small $\langle\alpha\rangle$ (see Fig. 5 (c)). Indeed, D -index is more suited to give a measure on the departure from the underlying spatial distribution than on spatial agglomeration. The spatial distribution of Singapore maritime industry is moderately agglomerated with many of its firms concentrated in the central business district. Because most of the firms in subsector 24 are located in unpopular districts, there is a departure from the underlying spatial distribution of all the firms in the maritime industry, and the results is a high D -index value. On the other hand, subsector 17 has been classified as highly agglomerated by $\langle\alpha\rangle$ but not by the D -index. Specifically, $D = 1.04$ for subsector 17. Interestingly, the small D -index value implies that firms in subsector 17 agglomerate at a popular location. In this case, popularity of the central business district serves as a natural advantage which induces an agglomeration of firms of subsector 17.

B. Spatial Co-occurrence

We calculate pairwise co-occurrences for the 34 maritime subsectors using 10 uniform grids with sizes that range from $500\text{ m} \times 500\text{ m}$ to $5000\text{ m} \times 5000\text{ m}$. Note that a random starting line (either vertical or horizontal and with a tolerance of 0 to 100 m) is assigned to begin the layout of each grid to avoid the influence of fixed boundaries. The averaged results (over the 10 grid sizes), $\langle C_{ij} \rangle$, then forms a 34×34 co-occurrence matrix. Hierarchical clustering is then performed on this co-occurrence matrix $\langle C_{ij} \rangle$ to produce the heatmap as shown in Fig. 6. Here, subsectors with a higher co-occurrence are arranged to situate near to each other.

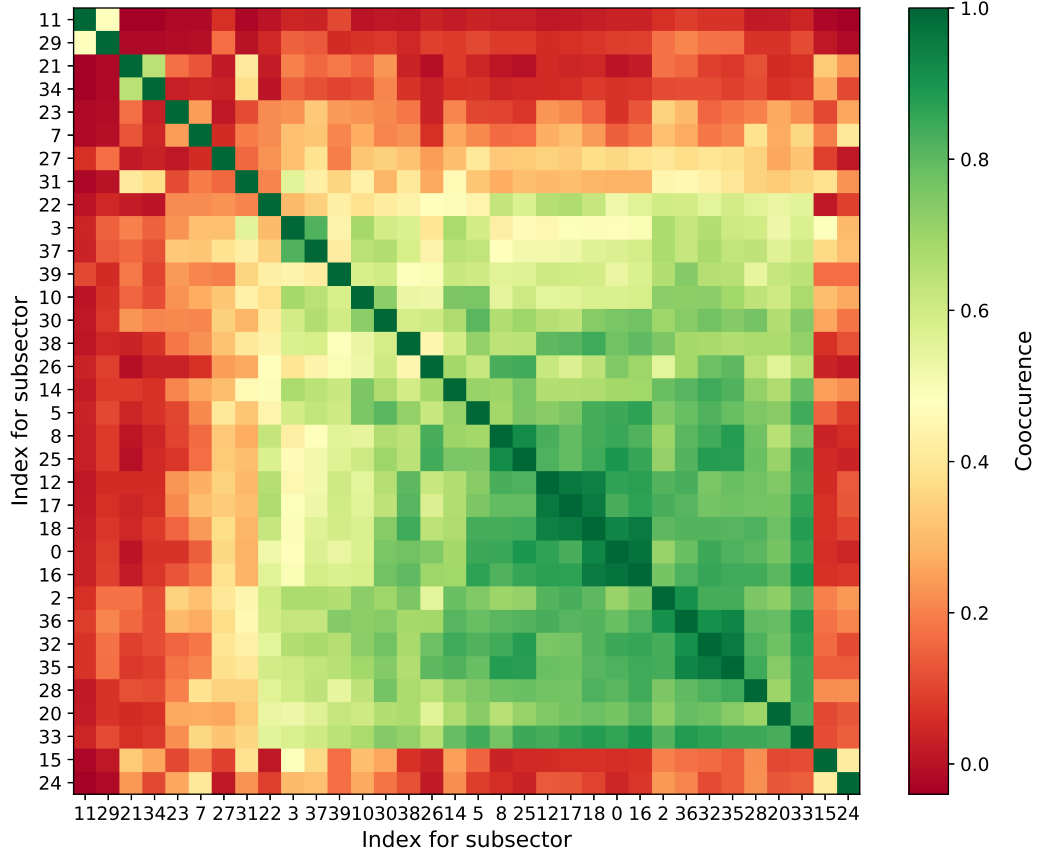


FIG. 6: Hierarchically-clustered heat map for the spatial co-occurrence matrix of the 34 subsectors in Singapore maritime industry under study.

We observe interesting cross subsector co-occurrences. Notably, subsectors 21 and 34 stand out as they do not co-occur with all of the other subsectors, but are having a moderate level of spatial co-occurrence (0.65) with each other. Spatial distribution of firms in these two subsectors is shown in Fig. 7 (a). Firms of both subsectors are found mainly in the south west and central of Singapore except for a number of type-34 firms which appear in the north of Singapore. Next, the heat map indicates that subsector 3 has a high spatial correlation (0.83) with subsector 37. The spatial distribution of firms in these two subsectors is shown in Fig. 7 (b), which shows clear spatial co-occurrence between the two subsectors. Spatial co-occurrence can also happen between three subsectors. For instance, subsectors

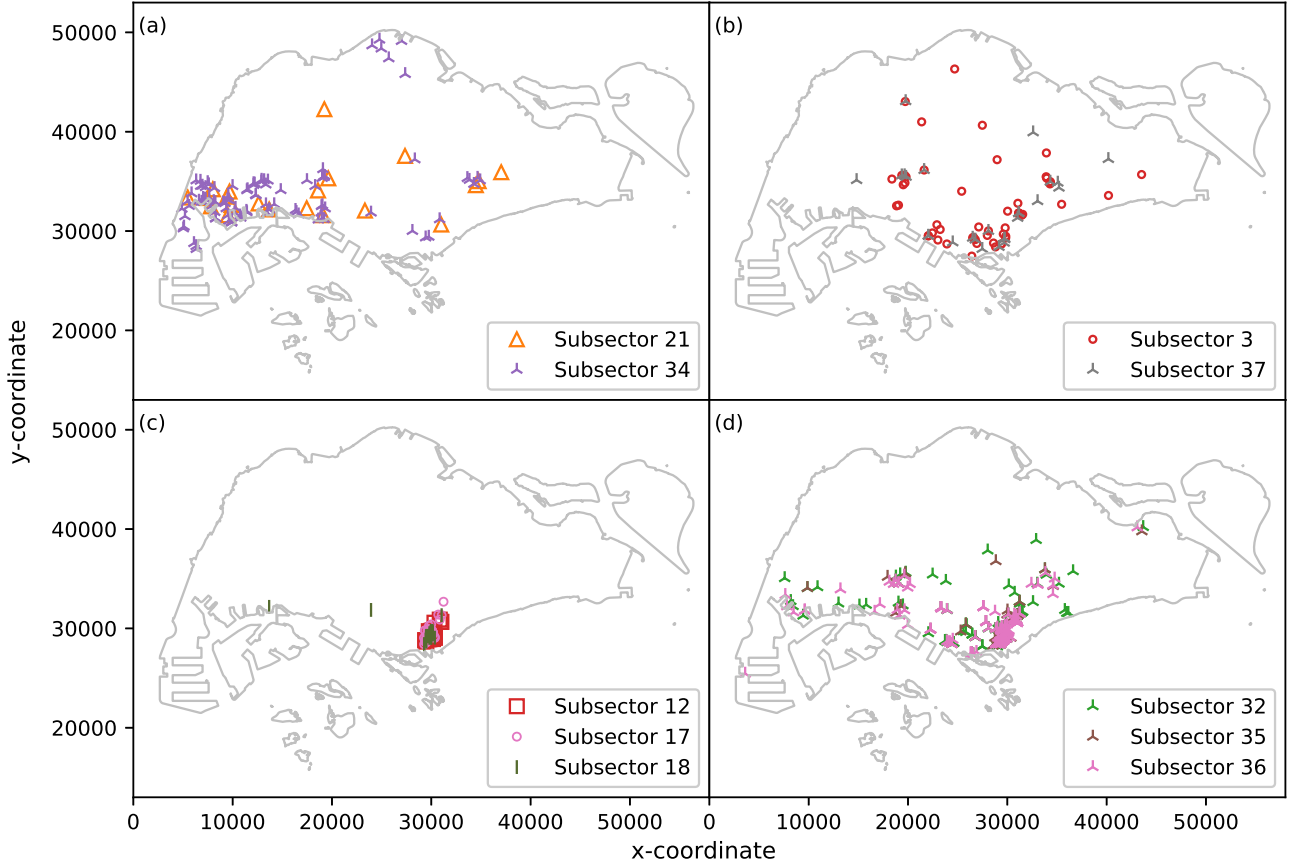


FIG. 7: The spatial distribution of firms of subsectors which co-occur with an average spatial co-occurrence of: (a) 0.65, (b) 0.83, (c) 0.95 and (d) 0.94.

12, 17 and 18 are observed to co-occur with each other as they cluster tightly at the central business district as shown in Fig. 7 (c). Their pairwise co-occurrences are in the range of 0.93 - 0.95. The spatial co-occurrence of three subsectors also occur at a larger length scale which is exhibited by subsectors 32, 35 and 36 as shown in Fig. 7 (d). The pairwise co-occurrences of these subsectors are in the range of 0.91 - 0.96.

It is informative to compare our defined spatial co-occurrence to the co-agglomeration index introduced by Ellison, Glaeser and Kerr [26], which is defined as follows;

$$\gamma_{ij} = \frac{\sum_{r=1}^R (s_{ri} - x_r)(s_{rj} - x_r)}{1 - \sum_{r=1}^R x_r^2}, \quad (9)$$

where i and j correspond to two different subsectors, and r indicates the geographic areas. s_{ri} is the fraction of subsector i 's firms that resides in area r and $\sum_r s_{ri} = 1$. $x_r = \frac{1}{M} \sum_{i=1}^M s_{ri}$, is the mean fraction of firms in region r across all M subsectors.

The interpretation of the strength of correlation coefficient depends on the context and purposes. A correlation coefficient of 0.7 is commonly recognized as strong [36] in the social sciences. Here, subsector i and j are observed to co-locate strongly when $\langle C_{ij} \rangle > 0.65$ (see for example the spatial distribution of subsectors 21 and 34 in Fig. 7

(a)). Thus, we compare between $\langle C_{ij} \rangle > 0.65$ and $\gamma_{ij} > 0$, which are conditions for the regime of co-occurrence or co-agglomeration respectively, to illustrate how these two measures can be in accord or different from each other. Note that this is a binary decision because not satisfying these conditions means absence of spatial co-occurrence or co-agglomeration. We observe a matching percentage of 62% between the two measures in this case. Note that the matching percentage can be different if we use a more moderate condition for spatial co-occurrence (i.e. $\langle C_{ij} \rangle > 0.5$) in our comparison. If we were to compare in terms of the non-binary form of Spearman rank correlation, we obtain a statistically significant correlation of 0.30 between co-agglomeration and spatial co-occurrence.

In summary, classifications via co-occurrence do not match well with classifications via co-agglomeration. One of the reasons for this mismatch is that co-agglomeration can give a negative value for a pair of subsectors with high positive spatial co-occurrence. Let us look at how this can happen from a simple example: $\hat{s}_i = \{0.72, 0.19, 0.09\}$, $\hat{s}_j = \{0.68, 0.21, 0.11\}$, and $x = \{0.7, 0.2, 0.1\}$. Although $\hat{s}_i$ and $\hat{s}_j$ has a high and positive spatial co-occurrence, co-agglomeration between them is negative because $s_{ri} > x_r$ when $s_{rj} < x_r$, and $s_{ri} < x_r$ when $s_{rj} > x_r$. Nonetheless, the two classifications are observed to typically match well for subsector pairs that highly co-occur with each other. In particular, all the highly co-occur subsector pairs (except subsectors 32 and 36) in Fig. 7 are classified as co-agglomerated. The reason for the mismatch between subsector 32 and 36 is due to the difference in sign between spatial co-occurrence and co-agglomeration as explained above. Let us illustrate what we mean. A small fraction of the firms in subsector 32 is found to locate at relatively unpopular regions. Because of which, the fraction of its firms in popular regions is slightly below the mean of these regions. On the contrary, firms of subsector 36 concentrate in popular regions and the fraction of its firms is larger than the mean of these regions. As a result, we obtain negative values for the product $(s_{ri} - x_r)(s_{rj} - x_r)$ in Eq. 9 for subsectors 32 and 36. Based on the same reasoning, it is clear that the products are also negative for these two subsectors in the unpopular regions. In consequence, $\langle C_{ij} \rangle$ gives a more transparent interpretation of spatial co-occurrence compared to γ_{ij} .

We next examine collocation of subsectors in the maritime industry using a distance-based measure whose method of evaluation is outlined in Ref. [37]. This measure of co-agglomeration, κ , is defined according to Eq. (7) with $\hat{K}$ and $\bar{K}$ given by the K -density, i.e. Eq. (6), where distances between firms is now measured between two subsectors instead of within a subsector. Similarly, we studied the K -density of such a pair of firms from two different subsectors up to $d = 2000$ m. The matching percentage is found to be 76% when $\langle C_{ij} \rangle > 0.7$ and $\kappa > 0$. The Spearman rank correlation, on the other hand, between the pairwise K -index and our spatial co-occurrence is statistically significant at 0.40. The rank correlation obtained is thus low, which implies an absent of a direct relationship between spatial co-occurrence and the K -density of the two subsectors. As an example, consider a subsector of 60 firms with 30 of its firms concentrate in the central business district (CBD) while the remaining 30 firms distribute over a large area, similar to how firms in subsector 3 distribute over Singapore. For simplicity, we assume that the firms which spread over a large area distribute equally in 5 other regions, i.e. 6 firms in each region. Then, the best way to site a second subsector of 30 firms such that the cross-subsector bilateral distance peaks at small distances in the K -density is to have all the 30 firms locate in the CBD. In this way, 900 (i.e. 30×30) pairs of firms from two different subsectors are

separated by short distances. On the other hand, if firms in the second subsector follows exactly the same pattern as how firms in the first subsector locate (i.e. 15 firms in CBD while the remaining 15 distribute equally in 5 other regions), only 540 (i.e. $15 \times 30 + 5 \times 3 \times 6$) of the pairwise distances are small. This explains why K -index classifies subsectors 3 and 37 as not co-agglomerated despite these subsectors having high spatial co-occurrence. Indeed, all pairs shown in Fig. 7 are classified as not co-agglomerated by the K -index because it is highly likely that randomly shuffled configurations possess a K -density that is skewed toward a smaller distance for the pairs concerned.

C. Following Index

Finally, in Fig. 8, we show the heat map of the average following indices of the 34 subsectors: $\langle F_{i \rightarrow j} \rangle = \frac{1}{10} \sum_{d=1}^{10} F_{i \rightarrow j}(200d)$. Here, each row gives the degree at which a subsector follows the rest of the subsectors of the maritime industry. For example, the first row gives the degree at which subsector 26 follows the other subsectors. Note that the subsectors are arranged from low to high coverage in the matrix. Unlike spatial co-occurrence, following indices are not symmetrical (see Fig. 8). For instance, $F_{5 \rightarrow 10} = 0.98$ while $F_{10 \rightarrow 5} = 0.49$, which implies that subsector 5 has a higher chance of following subsector 10. Such a following behaviour is significant because it occurs consistently throughout space (see Fig. 9), where firms of subsector 5 at different regions are always found to connect to firms of subsector 10 despite subsector 10 covers only about 23% of the space.

Subsectors 0, 8, 12, 16, 17, 18, 25, 26 and 38 appear to ‘follow’ most of the other subsectors in the industry. The high values in following index associated with these subsectors are however not significant because firms in these subsectors are observed to cluster at one region with a coverage of less than 5%. In contrast, subsectors 29 and 11 do not appear to follow any other subsectors in the industry. These two subsectors seem to operate independently of the other subsectors.

On the other hand, each column of the heat map quantifies the degree at which a subsector is being followed by other subsectors of the maritime industry. In particular, subsectors 2, 3, 10, 32, 33, 35 and 36 are followed by other subsectors with average following indices that are greater than 0.7. To determine whether these subsectors play a “leading” role in the “following behaviour” of the other subsectors, we take the coverage of these subsectors into account because a high value of “being followed” could result from the subsectors’ high coverage. Of particular interest is subsector 33 which covers only 12% of the space but is followed by other subsectors with an average following index of 0.71. Note that the coverage of the other six subsectors range from 17% to 30%. In addition, subsector 14 covers only 9% of the space but is followed by other subsectors with an average following index of 0.62. While how these subsectors lead the following behaviour of the other subsectors still need to be verified against the ground truth, an evaluation of the following index is expected to inform urban planners of potential leading subsectors in an industry.

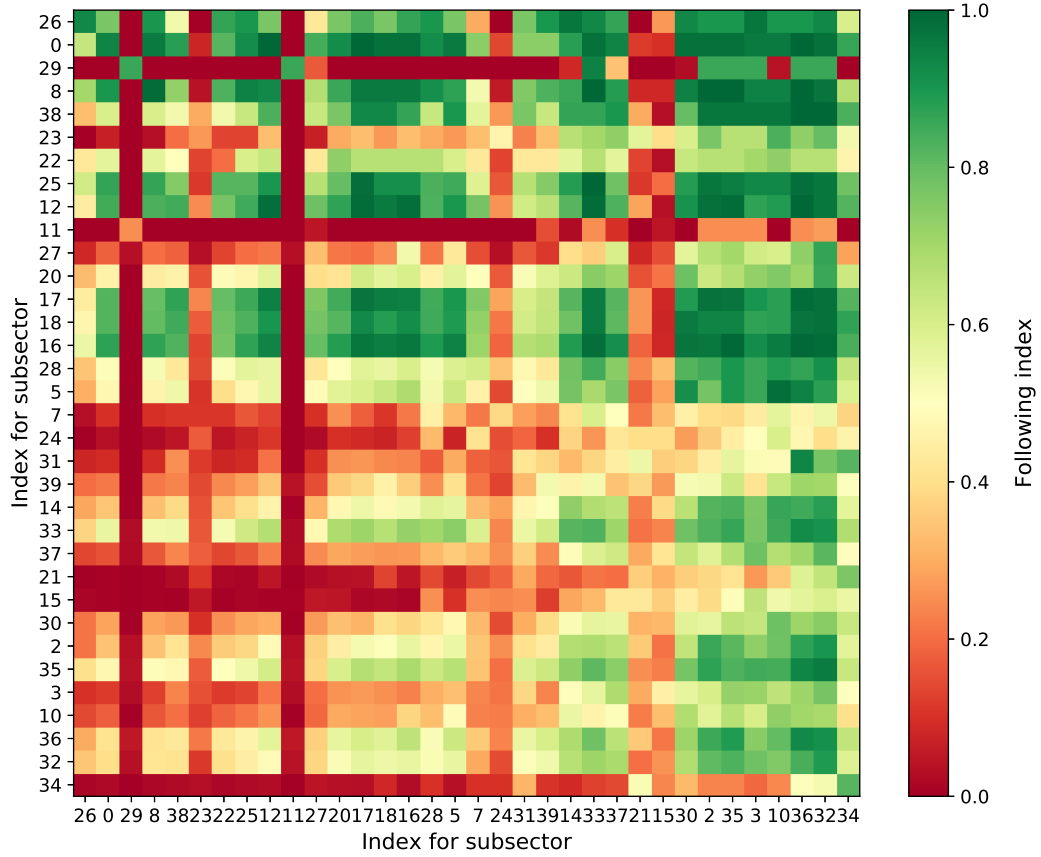


FIG. 8: Heat map for the matrix of following index for the 34 subsectors in Singapore maritime industry under study. Note that the subsectors are arranged from low coverage (top, left) to high coverage (bottom, right) in the matrix.

IV. AGGLOMERATION ECONOMY OF THE SINGAPORE MARITIME INDUSTRY

From Table S1, we observe that 82.3% of the subsectors of Singapore maritime industry has either the same or different type of firms (or even both types) spatially agglomerated with each other. While geometric concentration of firms is a central feature of agglomeration economics, we are interested to uncover the spatial clustering pattern of firms in the Singapore maritime industry to understand the underlying principles that give rise to this observed spatial distribution. Our hypothesis is that the maritime firms would choose to locate in a way that enables them to capture the positive impact from the localization externalities (such as access to (a) strategic information and technologies; (b) human resources with specialized technical, legal, and organization skills; (c) relevant financial resources and physical infrastructures; (d) lower transaction cost through scale economies etc) of the agglomeration economy.

Based on our clustering agglomeration and spatial co-occurrence indices, we observe the tendency of certain subsectors in the Singapore maritime industry to agglomerate and also co-locate at small length scales in the Central Business District (CBD) (see Fig. 5 (a)). These subsectors are: Finance, Law Firms, Law Firms (SCMA member), Accounting Services, Insurance, Media Service, Media and Public Relations, which are industrial sectors commonly found in the CBDs. These are advanced producer services (APS) that serve the maritime industry through ship

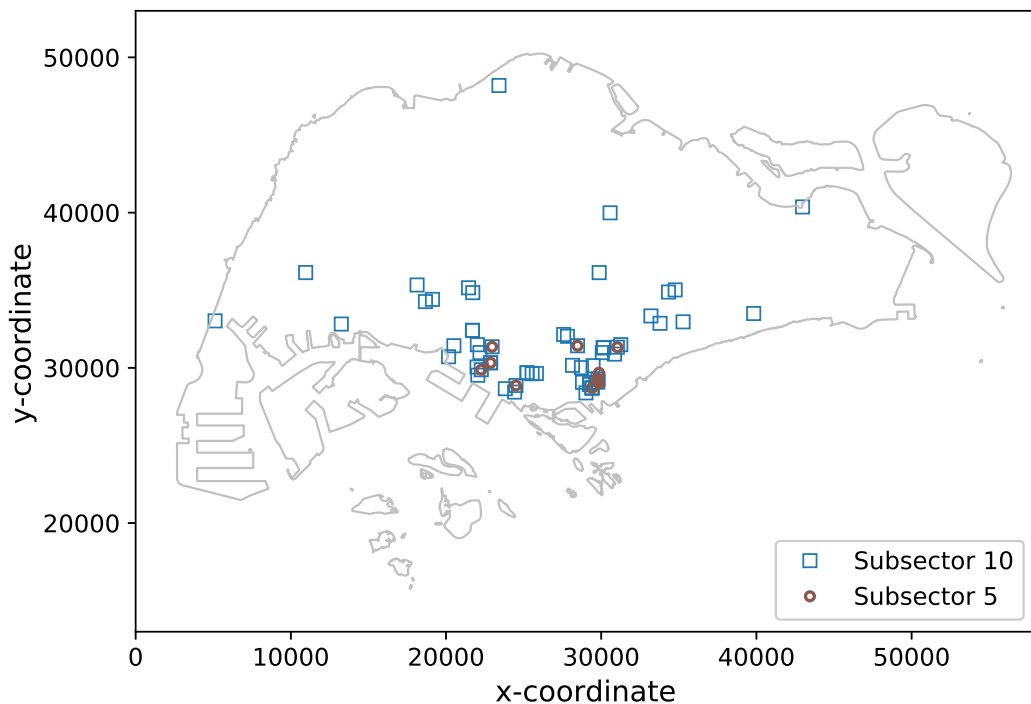


FIG. 9: The spatial distribution of firms in subsectors 5 and 10.

financing, ship and cargo insurance, legal representation, and key development advertisement [38]. The growth of these APS spatial cluster is a consequence of the fiscal policies legislated by the Singapore government (see Sec. I). Some of these APS may be global in operation, where they support the production functions of their clients regardless of their sector of operation. Together they form a global urban network of maritime APS with the network dominated by London, Singapore, and Hong Kong. By co-agglomerating in the CBD, these firms have access to a common pool of specialized labor, and through competing or collaborating with each other, gain new knowledge from the effects of spillover. In addition, we observe the subsector Cruise Operators to agglomerate and co-occur with the APS in the CBD, whose siting reflects a tendency to locate at a region that makes it prominent with easy access to a broad base of customers.

From our discussion in Section III A, spatial clustering can also occur at larger length scale, with the subsector “Ship Builder and Repairers” being the case in point (see Fig. 5 (b)). This clustering happens mainly in the South West which corresponds to sites occupied regionally with ship docking facilities and thus requires clear spatial separation between sites. There is also a cluster at the north with shipyard that has one of the deepest drydock in Southeast Asia, which possesses the capacity for naval and cruise vessels to dock without restrictions. Such clustering encourages competition between the firms which would foster productivity and economic growth as well as drive improvements in their process efficiency and product quality [39, 40].

We observe economic inter-dependency between maritime firms of different types when they locate close to each other. For example, Bunker Surveyors are found to highly co-occur with Ship Surveyors (see Fig. 7(b)). Similar to the APS firms in CBD, their co-location provides their clients with easy access to a suite of shipping services. In this case,

the services are surveying of bunkers and ships, which is necessary to certify the ship's sea-worthiness. While Bunker Surveyors co-agglomerate with Ship Surveyors, both subsectors have very low agglomeration with firms of their own type. Thus, we deduce that Bunker Surveyors and Ship Surveyors mainly derive economic gains through synergistic cooperation instead of competition in their exploitation of the localization externalities. On the other hand, Ship Agencies, Ship Management, and Ship Owner and Operators seem to optimise their returns from both economic collaboration and competition. This is inferred from their spatial co-occurrence (see Fig. 7(d)) and their clustering agglomeration (see Fig. 4). Because Singapore is a leading international maritime centre with many multinational shipping companies set up their ship ownership headquarters and ship management activities in Singapore, such clustering of subsectors could bring their economic cooperation and competition beyond the local regional scale to the international level. It is conceivable that the three subsectors have close economic dependence and work interaction with each other, with ship owners operating their ships through ship managements and ship agencies. Furthermore, the local agglomeration of each of the three subsectors leads to service provision at a lower cost through the external economies of scales [2].

The Singapore maritime industry traditionally begins with ship building and repairing activities. Subsequently, it has successfully transformed itself by diversifying into the offshore oil and gas construction and marine engineering services sectors. Eventually, Singapore has emerged as a world-leading offshore and marine engineering cluster [3]. This positive development results from the strategic cluster development policy instituted by the government and leads to the close business relationship between the Marine and Offshore Engineering subsector and the Ship Builders and Repairers subsector, which is exhibited by their high spatial co-occurrence even though Marine and Offshore Engineering subsector is dispersed in space (see Fig. 7(a)). In fact, Marine and Offshore Engineering subsector is observed to highly follow the Ship Builders and Repairers subsector (see Fig. 7(a)).

Following behaviour leverages on spatial proximity and density in agglomeration economy to site near firms-of-interest to extract economic benefits for its purposes. For example, Recruitment Agencies is found to follow the subsectors: Industry Organisations and Associations, Ship Management, Bunker Supplier, Ship Brokers, Ship Surveying, Research and Consultancy, for the provision of a pool of skilled manpower to this set of maritime subsectors. Interestingly, we observe strong following behaviour of Classification Societies to the subsectors of Education and Training as well as Research and Consultancy. The latter two subsectors could be the source of knowledge and manpower that support Classification Societies in its role of maintaining the technical standards on the construction and operation of ships and offshore structures.

V. DISCUSSION AND CONCLUSION

Spatial patterns could contain information on the economic features and relationship between subsectors of an industry. In this paper, we demonstrate the use of three quantitative measures to yield the spatial morphology

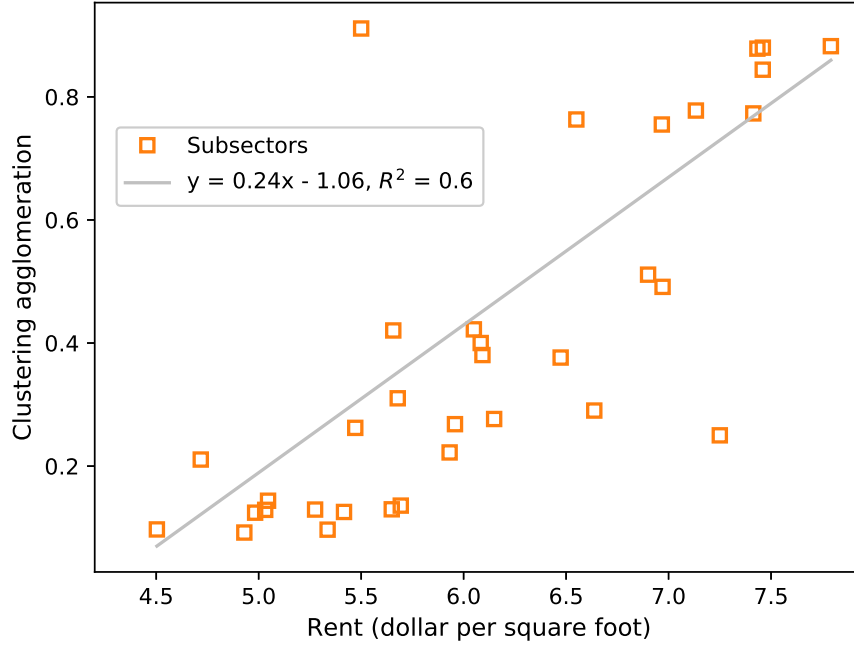


FIG. 10: A plot of the clustering agglomeration versus median rental rate [41] for the 34 subsectors of the Singapore maritime industry under study. Note that the clustering agglomeration has been averaged over a range of percolation distance from 200 m to 2000 m.

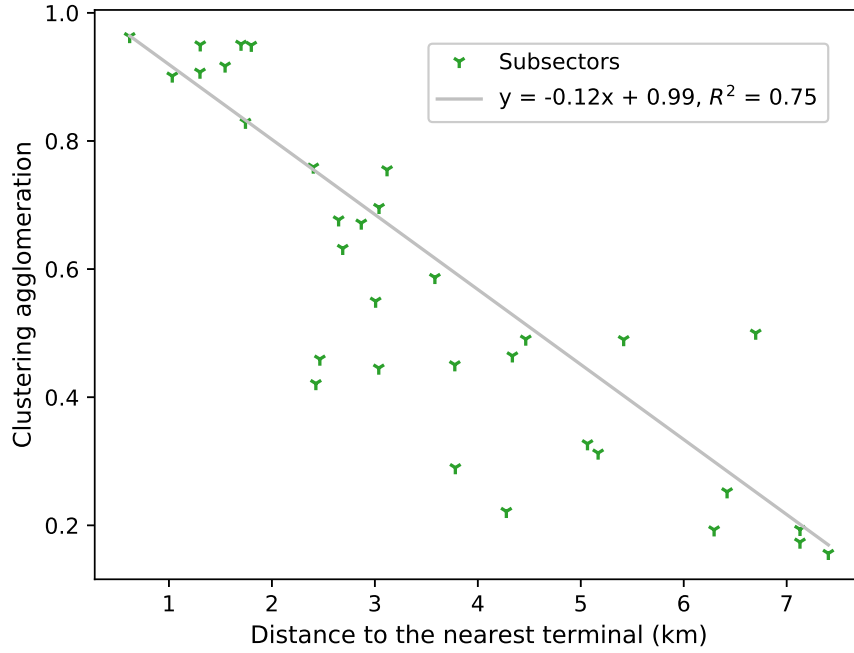


FIG. 11: A plot of the clustering agglomeration versus average distance to the nearest terminal for 33 subsectors in Singapore maritime industry (ferry terminals is excluded). The six shipping terminals used in the estimation of the nearest distance are Keppel Terminal, Tanjong Pagar Terminal, Pasir Panjang Terminal, Sembawang Terminal, Jurong Terminal and Tuas Terminal. The clustering agglomeration has been averaged over a range of percolation distance from 200 m to 5000 m.

of firms in terms of its agglomeration, their spatial relationship with each other, and also the potential presence of leading and following behaviour. From these measures, we are able to draw insights into certain economic functionality and dependency of the subsectors in the Singapore maritime industry. Such an evaluation would become especially pertinent when the spatial data becomes more complex with firms from multiple industries and with the inclusion of additional commercial bodies. Furthermore, the simplicity of our measures would enable computation from such large datasets efficient. Nonetheless, verification from ground truth is necessary as spatial correlation does not imply the presence of causal relationship between the subsectors or the following behaviour. Future studies could include additional attributes like size of firms, status of firms (such as multi-national, domestic, headquarter), and other specific characteristics of maritime firms, to gain deeper insights on the agglomeration economy of the Singapore Maritime Industry.

We have also compared our measures of clustering agglomeration and spatial co-occurrence against the conventional agglomeration indices of K -index, D -index, and EGK co-agglomeration index. Unlike these indices from the literature which were developed through the provision of a statistical testing framework or from a location choice model, our measures are more direct in capturing the spatial morphological features and do not need employment data like the EGK co-agglomeration index. (Note that in order to make the comparison, we have adapted EGK co-agglomeration index to the use of spatial morphological data). Like the K -index, the clustering agglomeration is a continuous space index. On the contrary, spatial co-occurrence and following index are essentially discrete space indices (like the D -index and the EGK co-agglomeration index) although we have removed their discrete space dependency through averaging over different grid cell size. Our results show that there is some level of correspondence between the estimates from our measure of clustering agglomeration and the K -index, while we have found substantial deviation between estimates from our indices against that from the D -index and the EGK co-agglomeration index due to subtle differences in definition.

Moreover, we have shown that our combined morphological measures of clustering agglomeration, spatial co-occurrence and following index can serve as tools for the testing of hypothesis in an agglomeration economy. We have demonstrated this role of the measures by uncovering whether firms in the Singapore maritime industry exploit the benefits of localization externalities in the agglomeration economy to site their operation. From empirical analysis through these measures, we observe that maritime firms indeed have a tendency to agglomerate. We surmise that their basis of co-agglomeration is to engage on its positive effects, which are a pooling of customer base and skilled labor sources, tapping on the gains of knowledge spillover, and reaping the benefits of external economy of scale. Specific subsectors also adopt synergistic collaboration to exploit the advantages of localization externalities, or to engage in competition to enhance their operational efficiency and productivity through the agglomeration economy. In particular, we would like to emphasize the importance of government policies in engaging the agglomeration economy. In the case of Singapore, policy had been applied to attract the establishment of anchor maritime firms through the release of subsidized land parcels. New entrants to these land parcels then leveraged on the expertise of these early entrants and grew organically, which resulted in the present cluster agglomeration. Here, our morphological measures become

useful. Through examining and elucidating the behaviour of these new entrants with our measures, the government could then institute appropriate policies to spatially distribute firm-types in future allocation in a self-organized way to optimize land-use efficiency. With the continual development of land and infrastructures by the Singapore government for the industry, the occurrence of diseconomies of agglomeration has been largely circumvented.

In addition to the morphological dimension, we have examined into other factors that account for the spatial pattern of agglomeration. One factor is rent where we observe the presence of a positive relationship: a higher rate of rental of a location is associated with a higher spatial agglomeration of maritime firms (see Fig. 10). This is in line with the literature where it was found that greater productivity level which occurs at a higher-observed measure of agglomeration economies will act to increase the relative land rent of that region [42–44]. The positive agglomeration elasticity of rent reveals the presence of external economies, and we found that the highest level happens at the CBD from the APS. Another factor is distance to port infrastructure, where we observe a decrease in the agglomeration of the maritime firms with an increase in this factor (see Fig. 11). From the perspective of agglomeration economy, this can be understood from the fact that spatial proximity to the port infrastructure would contribute positively to the productivity of the maritime firms, which explains the observed negative relationship [2, 45]. These additional analysis results on other factors have further affirmed our hypothesis that the Singapore Maritime Industry is engaged in the agglomeration economy.

-
- [1] Monteiro, P., de Noronha, T. & Neto, P. (2013). A differentiation framework for maritime clusters: comparisons across Europe. *Sustainability*, 5, 4076–4105.
 - [2] Talley, W. K. (2009). *Port Economics*, New York:Routledge.
 - [3] Wong, P. K., Ho, Y. P. & Singh, A. (2010). *Industrial cluster development and innovation in Singapore*, Institute of Developing Economics (IDE-JETRO), Palgrave-Macmillan, pp. 50 – 116.
 - [4] Marshall, A. (1920). *Principles of Economics*. London: MacMillan.
 - [5] Jaffe, A. B., Trajtenberg, M. & Henderson, R. (1993). Geographic localisation of knowledge spillovers as evidence by patent citations. *Quarterly journal of Economics*, 108(3), 577–598.
 - [6] Doeringer, P. B. & Terkla, D. G. (1995). Business strategy and cross-industry clusters. *Economic development quarterly*, 9, 225–237.
 - [7] Breschi, S., & Lissoni, F. (2009). Mobility of skilled workers and co-invention networks: An anatomy of localized knowledge flows. *Journal of Economic Geography*, 16(2), 1–30.
 - [8] Porter, M. E. (2000). Location, competition, and economic development: Local clusters in a global economy. *Economic development quarterly*, 14(1), 15–34.
 - [9] Parr, J. B. (2002). Agglomeration economies: Ambiguities and confusions. *Environment and Planning A*, 34(4), 717–731.
 - [10] Duranton, G. & Puga, D. (2004). Micro-foundations of urban agglomeration economies. *Handbook of regional and urban economics*, 4, 2063–2117.

- [11] Flores, M., Villarreal, A. & Flores, S. (2017). Spatial Co-location Patterns of Aerospace Industry Firms in Mexico. *Appl. Spatial Analysis*, 10, 233–251.
- [12] Arcaute, E., Molinero, C., Hatna, E., Murcio, R., Vargas-Ruiz, C., Masucci, A. P. & Batty, M. (2016). Cities and regions in Britain through hierarchical percolation. *Royal Society Open Science* 3, 150691.
- [13] Cottineau, C., Finance, O., Hatna, E., Arcaute, E. & Batty, M. (2019). Defining urban clusters to detect agglomeration economies. *Environment and Planning B: Urban Analytics and City Science*, 46(9), 1611–1626.
- [14] Rigby, D. L. & Brown, W. M. (2015). Who benefits from agglomeration? *Regional Studies*, 49, 28 – 43.
- [15] Zheng, X.-P. (2007). Economies of network, urban agglomeration, and regional development: a theoretical model and empirical evidence. *Regional Studies*, 41, 559 – 569.
- [16] Bettencourt, L. M. A., Lobo, J., Helbing, D. and West, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences*, 104, 7301–7306.
- [17] Cottineau, C., Hatna, E., Arcaute, E., & Batty, M. (2017). Diverse cities or the systematic paradox of urban scaling laws. *Computers, Environment and Urban Systems*, 63, 80–94.
- [18] Merk, O. (2013). The competitiveness of global port cities: synthesis report. *OECD Regional Development Working Papers*, No. 2013/13, OECD Publishing, Paris.
- [19] Hervás-Oliver, J.-L., Sempere-Ripoll, F., Alvarado, R. R. & Estelles-Miguel S. (2018). Agglomerations and firm performance: who benefits and how much? *Regional Studies*, 52, 338 –349.
- [20] Han, W., Zhang, Y., Cai, J., & Ma, E. (2019). Does urban industrial agglomeration lead to the improvement of land use efficiency in China? An empirical study from a spatial perspective. *Sustainability*, 11, 986.
- [21] Chen, Y., Chen, X., Liu, Z., & Li, X. (2020). Understanding the spatial organization of urban functions based on collocation patterns mining: a comparative analysis of 25 Chinese cities. *Cities*, 97, 102563.
- [22] Ellison, G. & Glaeser, E. L. (1997). Geographic concentration in U.S. manufacturing industries: a dartboard approach. *J. Polit. Econ.*, 105, 889–927.
- [23] Maurel, F. & Sédillot, B. (1999). A measure of the geographic concentration in French manufacturing industries. *Regional Science and Urban Economics*, 29, 575–604.
- [24] Rysman, M. & Greenstein, S. (2005). Testing for agglomeration and dispersion. *Economics Letters*, 86, 405–411.
- [25] Mori, T., Nishikimi, K. & Smith, T. E. (2005). A Divergence Statistic for Industrial Localization. *The Review of Economics and Statistics*, 87, 635–651.
- [26] Ellison, G., Glaeser, E. L. & Kerr, W. R. (2010). What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns. *American Economic Review*, 100, 1195–1213.
- [27] Duranton, G. & Overman, H. G. (2005). Testing for Localization Using Micro-Geographic Data. *The Review of Economic Studies*, 72, 1077–1106.
- [28] Marcon, E. & Puech, F. (2003). Evaluating the geographic concentration of industries using distance-based methods. *Journal of Economic Geography*, 3, 409–428.
- [29] Data for Singapore maritime firms is obtained from the Singapore maritime services guide at the following website: <https://www.timesdirectories.com/ebook/smsg/2017/mobile/index.html>.
- [30] Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, AAAI*

Press, 226–231.

- [31] Huynh, H. N., Makarov, E., Legara, E. F., Monterola, C. & Chew, L. Y. (2018). Characterisation and comparison of spatial patterns in urban systems: A case study of U.S. cities. *Journal of Computational Science*, 24, 34–43.
- [32] Alves, L. G. A., Lenzi, E. K., Mendes, R. S. & Ribeiro, H. V. (2015). Spatial correlations, clustering and percolation-like transitions in homicide crime. *Europhysics Letters*, 111, 18002.
- [33] Bertinelli, L. & Decrop, J. (2005). Geographical agglomeration: Ellison and Glaeser’s index applied to the case of Belgian manufacturing industry. *Regional Studies*, 39, 567 – 583.
- [34] Hidalgo, C. A., Castañer, E. & Sevtsuke, A. (2020). The amenity mix of urban neighborhoods. *Habitat International*, 106, 102205.
- [35] Leslie, T. & Kronenfeld, B. (2011). The colocation quotient: a new measure of spatial association between categorical subsets of points. *Geographical Analysis*, 43, 306–326.
- [36] Akoglu, H. (2018). User’s guide to correlation coefficients. *Turk. J. Emerg. Med.*, 18, 91–93.
- [37] Duranton, G. & Overman, H. G. (2008). Exploring the detailed location patterns of U.K. manufacturing industries using microgeographic data. *Journal of Regional Science*, 48, 213–243.
- [38] Jacobs, W., Koster, H. & Hall, P. (2011). The location and global network structure of maritime advanced producer services. *Urban Studies*, 48, 2749 – 2769.
- [39] Porter, M. E. (1990). The competitive advantage of nations. *Harvard Business Review*, 68, 73–93.
- [40] Hylton, P. J. & Ross, C. L. (2018). Agglomeration economies’ influence on logistic clusters’ growth and competitiveness. *Regional Studies*, 52, 350 –361.
- [41] Data for average rent is obtained from the website: <https://officefinder.com.sg/trend/singapore-office-rental-rates-guide-by-district/>.
- [42] Dekle, R. & Eaton J. (1999). Agglomeration and land rents: evidence from the prefectures. *Journal of Urban Economics*, 46, 200 – 214.
- [43] Koster, H. R. A., van Ommeren, J. & Rietveld, P. (2014). Agglomeration economies and productivity: a structural estimation approach using commercial rents. *Economica*, 81, 63 – 85.
- [44] Drennan, M. P. & Kelly, H. F. (2011). Measuring urban agglomeration economies with office rent. *Journal of Economic Geography*, 11, 481 – 507.
- [45] Jacobs, W., Koster, H. R. A. & van Oort, F. (2014). Co-agglomeration of knowledge-intensive business services and multinational enterprises. *Journal of Economic Geography*, 14, 443 – 475.

Supplementary Information: A data analytic elucidation on the spatial agglomeration of Singapore maritime industry

A. Results for other indexes

In this section, we show the results of K -index, D -index, EGK co-agglomeration index and K co-agglomeration index for subsectors in Singapore maritime industry. First, the K -indices for the 34 subsectors are given in Fig. S1, where the K -density was studied over a specified distance horizon. We vary this distance horizon over two ranges of length scale: 200 m to 2000 m, and 200 m to 5000 m. We observe similar results for these two cases (see Fig. S1). Based on the classification via K -index, we determine that each of the following subsectors 2, 32, 36, 35, 33, 8, 25, 17, 18, 12, 16, 0 and 26 agglomerates at a distance smaller than 2000 m, while each of the subsectors 10, 34, 23, 14, 3 and 38 agglomerates at a distance larger than 2000 m.

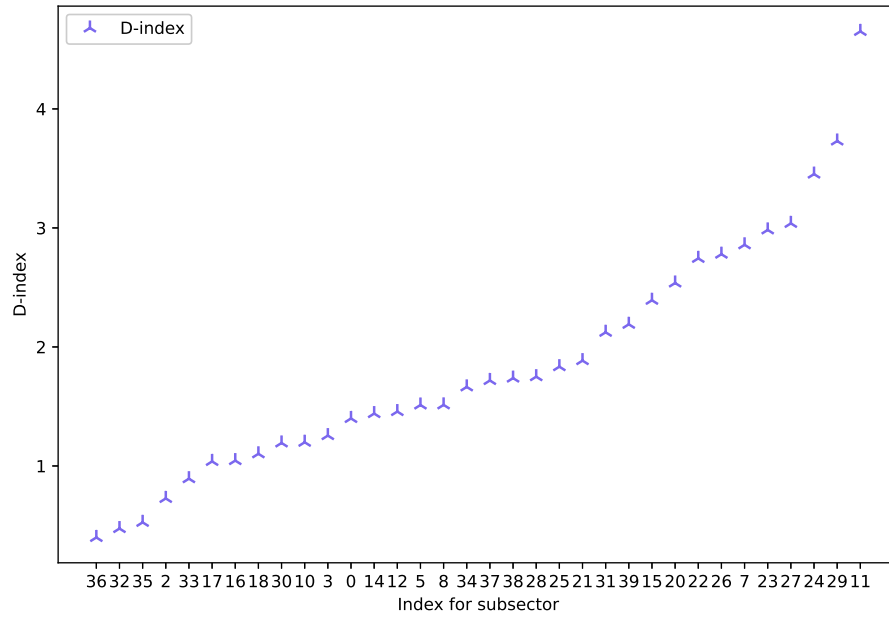
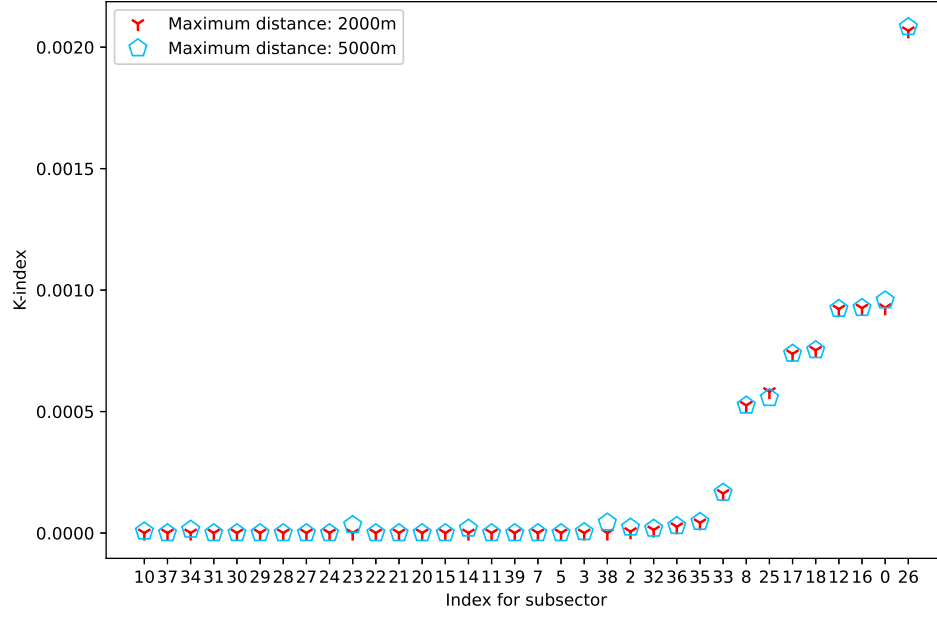
Figure S2 illustrates the application of D -indices to the subsectors of Singapore maritime industry. Here, we adopt the subzone boundary provided by Singapore's Master Plan 2014 (published at <https://data.gov.sg/dataset/master-plan-2014-subzone-boundary-no-sea>) as our geographic division. The resultant D -indices are found to range from 0 to 5. Ten subsectors (i.e. 11, 29, 24, 27, 23, 7, 26, 22, 20 and 15) possess D -indices which are larger than a threshold of 2.6. Among these subsectors, only subsectors 29 and 26 are classified as highly agglomerated by our clustering agglomeration index (refer to Fig. 4 in the main text).

In Fig. S3, we present the hierarchically-clustered heat map of the EGK-index which measures the degree of co-agglomeration among Singapore maritime industry's subsectors. Surprisingly, we observe high co-agglomeration between subsectors 29 and 11 with the EGK-index, which differs from our outcome using spatial co-occurrence. The reason for the high co-agglomeration results from firms in both of these subsectors distributing only in a few regions and a substantial fraction of them locating in a less popular region. The effect is a positive contribution to γ_{ij} . Although the distributions of these firms at four other unpopular regions are not consistent, the negative contributions are not significant as one of the subsectors has a zero s_{ri} (magnitude of $(s_{ri} - x_r)$ is thus small) when the other has a nonzero s_{rj} . Furthermore, contributions from the rest of the regions are positive because both subsectors have no firms in these regions. It is interesting to note that there exists only a moderate correlation between subsectors 29 and 11 if the analysis were made based on spatial co-occurrence (see Fig. 6 in the main text).

Finally, we show in Fig. S4 the co-agglomeration probed by K -index for two distance ranges for the Singapore maritime industry.

B. Details on agglomeration, co-occurrence, and following behavior for each subsector

In Table S1, we show the agglomeration, co-occurrence, and following behavior for each subsector. Out of the 34 subsectors we studied, 14 subsectors were found to agglomerate and co-agglomerate, 14 subsectors were found to either agglomerate or co-agglomerate, and 6 subsectors were found to neither agglomerate nor co-agglomerate. Furthermore,



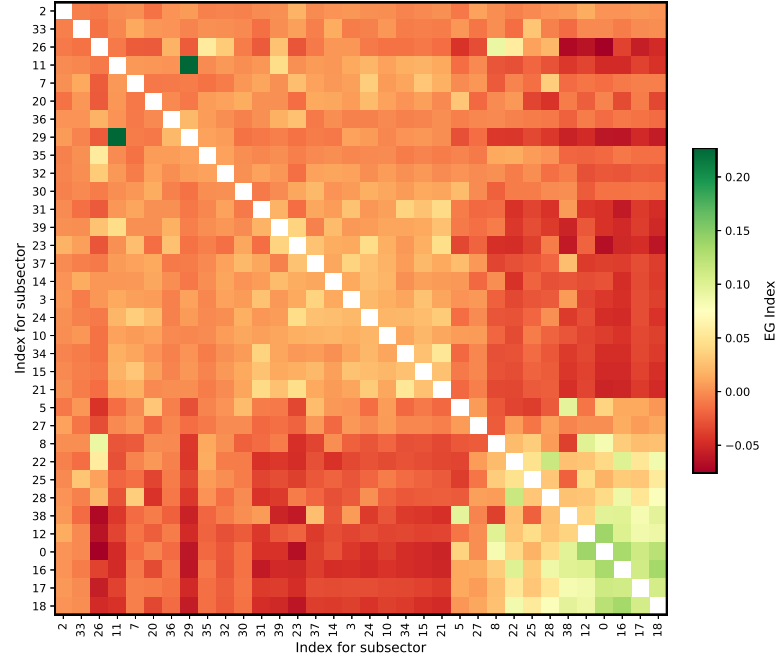


FIG. S3: Hierarchically-clustered heat map of the EGK co-agglomeration indices for the 34 subsectors of Singapore maritime industry. Note that the diagonal elements are irrelevant and much larger than all the other elements. We have thus masked these entries.

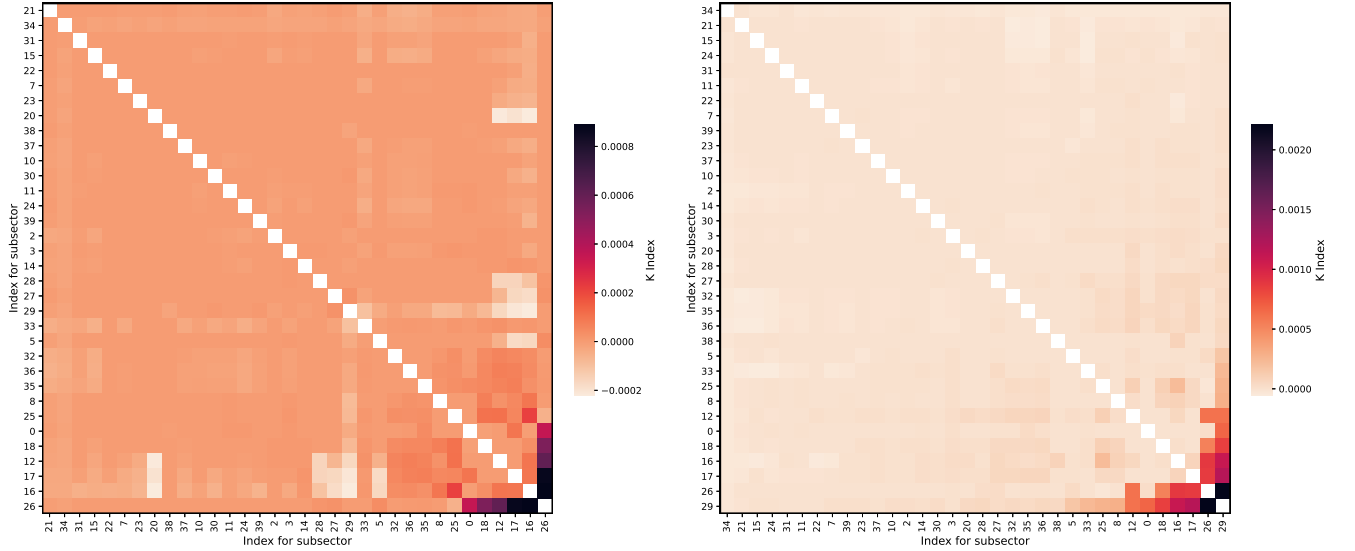


FIG. S4: Hierarchically-clustered heat maps of co-agglomeration examined by the K -indices for the 34 subsectors of Singapore maritime industry with maximum distance $d_{max} = 2000$ m (left) and $d_{max} = 5000$ m (right). Note that the diagonal elements are irrelevant and are thus masked without any entries.

14 subsectors were found to exhibit significant following behavior.

In Fig. S5, we show the co-occurrence network of the Singapore maritime industry. The co-occurrence network was constructed based on the heat map shown in Fig. 6, where pairs of subsector which have a co-occurrence greater than 0.65 are connected by links in the network. The co-occurrence network provides a visualization of the overall structure of spatial co-occurrence of maritime subsectors within the industry.

Index	Subsector	Agglomerate	Co-occur	Following
0	Accounting services	✓	✓	✗
1	Arbitration centre	-	-	-
2	Bunker suppliers	✗	✓	✓
3	Bunker Surveyors	✗	✓	✗
4	Claims Agency (SCMA members)	-	-	-
5	Classification societies	✗	✓	✓
6	Coal trading	-	-	-
7	Communication Services	✗	✗	✗
8	Cruise operators	✓	✓	✗
9	Cruise terminal	-	-	-
10	Education and training	✗	✓	✗
11	Ferry terminals	✗	✗	✗
12	Finance	✓	✓	✗
13	Freight derivatives clearing	-	-	-
14	Industry organisations and associations	✗	✓	✓
15	Inspection services	✗	✗	✗
16	Insurance	✓	✓	✓
17	Law firms	✓	✓	✓
18	Law firms (SCMA members)	✓	✓	✓
19	Legal services	-	-	-
20	Loss adjusting	✗	✓	✗
21	Marine and offshore engineering	✗	✓	✓
22	Marine consultancy	✓	✓	✗
23	Marine equipment and supplies	✓	✗	✗
24	Marine security	✗	✗	✗
25	Media and public relations	✓	✓	✗
26	Media services	✓	✓	✗
27	Port operators and representatives	✗	✗	✗
28	Recruitment agencies	✗	✓	✓
29	Regional ferry operators	✓	✗	✗
30	Research and consultancy	✗	✓	✓
31	Salvage	✗	✗	✓
32	Ship agencies	✓	✓	✗
33	Ship brokers	✓	✓	✓
34	Ship builder and repairers	✗	✓	✗
35	Ship management	✓	✓	✓
36	Ship owners and operators	✓	✓	✗
37	Ship surveying	✗	✓	✓
38	Shipowners - bunker tankers	✓	✓	✗
39	Shipping logistics	✗	✓	✓

TABLE S1: Maritime industry subsectors which agglomerate, co-occur with other subsector(s), or following other subsector(s). Here, $\alpha = 0.36$ is used as a threshold to determine whether a subsector agglomerates or not. The threshold for co-occurrence is set at $\langle C_{ij} \rangle > 0.65$ and a subsector is given a tick under “Co-occur” if it co-locates with at least one other subsector. The threshold for following index is set at 0.7 and a subsector is considered to exhibit following behavior if it follows at least one other subsector which has a coverage of 0.2 or smaller.

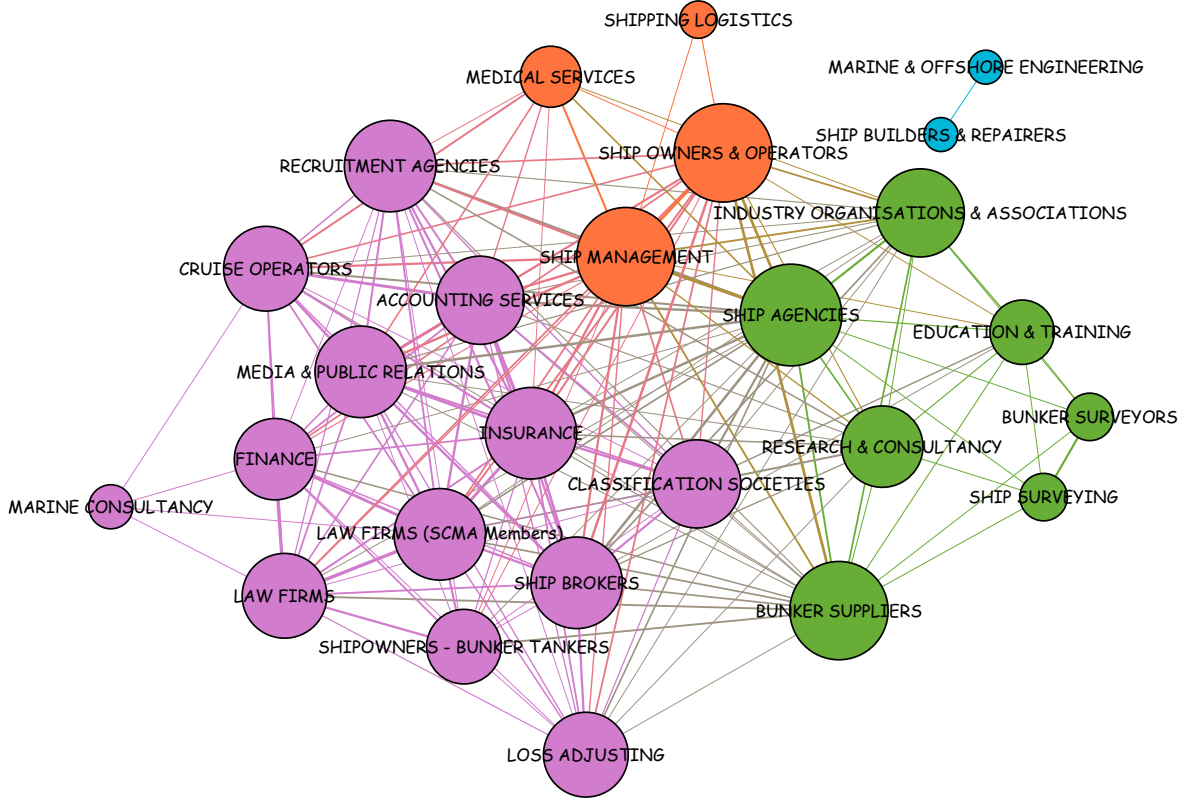


FIG. S5: Co-occurrence network for subsectors which co-occur with at least one other subsector with $\langle C_{ij} \rangle > 0.65$. Here, size of a node (i.e. a subsector) is proportional to degree (i.e. number of connection) of the node. In addition, nodes in the same community are more densely connected with each other while nodes in different communities are less connected. Nodes in the same community are represented by the same color. Lastly, thickness of an edge is proportional to the co-occurrence between the two subsectors.

C. Potential practical issue for the calculation of agglomeration

It is extremely inefficient to determine clusters by means of the DBSCAN algorithm for large datasets. To improve efficiency, we grid the data points before performing DBSCAN. Specifically, we represent the data points within a cell in the grid by one of the data points which was selected randomly. Cluster analysis is then performed through these randomly selected data points. This procedure reduces the number of data points used to perform DBSCAN, thus speed up the calculation of clusters. We also note down the number of data points in each cell in the grid. All data points in a cell shall belong to the same cluster as the representative data point of the cell. To find the frequency of a cluster, we add up the frequencies of all the points in the cluster. Agglomeration calculated through this method is accurate as long as the grid size is much smaller than the percolation distance. For example, with a grid size of 50 m, we are able to reproduce the average agglomeration of the 34 Singapore maritime subsectors with a percolation distance that ranges from 200 m to 2000 m.