Business Agglomeration-Based Decision Support Systems to Identify Prospective Locations for New Businesses

Harya Widiputra, Marsudi Kisworo, Agnes Novita, Tiolina Evi
Faculty of Information Technology
Asian Banking Finance and Informatics Institute Perbanas
Jakarta, Indonesia
e-mail: harya;marsudi;agnes.novita;tiolina@perbanasinstitute.ac.id

***Abstract*—Selecting the right location when establishing new business firm is one imperative key to a successful growth of an establishment. Additionally, previous studies have also found that business firms form business agglomerations that enable these enterprises to collaborate. However, this agglomeration also produces some latent threats, for instance the intraspecific competition between establishments belongs to the same group. Thus, it is then logical to consider the task of selecting business location for a new establishment as a mission of identifying prospective business agglomeration in which the new establishment would be able to compete with existing business firms. This study develops a decision support system that helps to recognize prospective locations for new businesses by incorporating the competition indices within existing business agglomerations. Results from conducted experiment suggest that the developed system is capable to complete such task with a reasonable degree of acceptance.**

Keywords: business agglomeration, business location, competition indices, decision support system.

I. Introduction

Previous studies have found that business units are not spread out independently and also not being established exclusively. Conversely, people tend to start their business near to other existing business units and then directly or indirectly construct business agglomerations [1]. These groups of business units are usually formed due to the closeness of geographic position and also similarity of business nature. The existence of these agglomerations brings not only positive effect, which enables business units to collaborate, but also negative effect known as the intraspecific competition that might have significant influence to the development of competing enterprises [2,3].

Nevertheless, location has also always been one of the key factors that determine the success of a business firm, both for large companies and small or medium enterprises [4]. Good selection of business site would definitely have great influence to the development of the company in the future. However, the definition of location in this case is not limited to a particular address neither a precise position in terms of geographical location defined by a set of latitude or longitude. Yet, location here relates more to a certain business agglomeration in a geographical region. Opening new business in an unpromising cluster might results in a competition that cannot be won due to the discriminating intraspecific competition factor between business units of a particular agglomeration.

Nevertheless, the commonly used analyses for business site selection are not usually taking into account the intraspecific competition in each agglomeration when making final selection. Consequently, a business location selection is made by making evaluation to the availability of some industrial components which are vital to the company, i.e. the existence of raw materials, labor, access to other required resources, etc. [5]

Contrariwise, previous study suggests that a better selection for business site can be made if one is capable to identify the business agglomerations within a particular geographical area [1, 2]. By being able to do so, maps of intraspecific competition in each business clusters can then be developed. Using these maps of competition as a basis, future condition of a new business establishment can be predicted for each cluster. The best location for the new business unit should then be the agglomeration in which the new establishment is predicted to perform the best.

This research develops a graph-clustering algorithm, which is based on previous study by [6], to identify business agglomerations in a particular area and then to apply such knowledge to identify prospective location for a new business firm. Correspondingly, the main objective of this study is to develop a decision support system that helps making selection of new business location. In addition, reviews of related studies are outlined in the upcoming section whilst designs of the developed algorithms are described in Section 3. Section 4 outlines some discussion in relation to results from conducted experiments and finally specific conclusions and future works are given in Section 5.

II. Related Works

As it was rationalized in previous section, business firms in a geographical area do not subsist individually yet they either directly or indirectly constitute business agglomerations. This actuality leads to the perception that choosing prospective location for new business relates to the identification of business agglomeration with acceptable level of competition hence the new establishment would be able to outlast or even win the business competition. This section outlines previous related works required to develop a DSS for selecting prospective business location for a new establishment.

A. Graph-Clustering

Graph-clustering practice is a process to segment a particular graph that results in a number of agglomerations consisting of nodes belongs to the graph. As in the basic concept of clustering process, intra-distance between nodes belong to the same cluster should be smaller than inter-distance between nodes in different clusters.

A significant difference between a graph-clustering process and the common clustering process lays in the nature of how the clusters are constructed. In a common clustering process, objects are grouped based on the similarity between attributes that define the objects. Contrary, in graph-clustering nodes are put into the same cluster by assessing the connectivity and also structural similarity between them [7].

Some other graph-clustering algorithms segment a single graph into clusters of nodes based on distance in the space domain and also linkage between the nodes [8].

In addition, [9] proposed a graph-clustering technique that consists of two clustering process. The first level of clustering process is to group nodes with their nearest neighbor to construct the initial clusters. Once these initial clusters are identified to process is continue by merging these clusters based on the similarity of affinity indices between clusters. The developed DSS in this study employs this concept whereas the competition level between business firms serves as the affinity indices.

B. Semi-supervised Learning for Classification Task

Semi-supervised learning for classification task is simply giving labels to unlabeled samples after a learning process using some labeled training set has been conducted. This process usually involves a classifier, i.e. Naïve Bayes, a neural network, a decision tree model, etc. [10, 11, 12]

[6] in their work suggested that the task of predicting upcoming performance of a new business firms could be transformed into a semi-supervised classification task. A key requirement to completing such task is to convert the performance indicator of a business firm, which is usually a continuous value represented by number of sales, revenue or profit, into some discrete values of distinct categories. In addition, information about the quality factors and their value that are assumed to have significant influences to the performance of an establishment is compulsory since they will serve as the input attributes of the classification task. Consequently, the predicted class label would be the business performance category of a new business firms.

III. Business Agglomeration-Based DSS

This research aims to build a decision support system that helps in choosing the most reasonable location, that is a business agglomeration, for a new business unit given that the quality vector of this new firm in known. The DSS that is to be built is inspired by a previous work that transformed the problem of selecting new business location into a semi-supervised learning task [6].

In general the proposed DSS in this study applies analogous steps as in the work by [6]. Core methods of the developed DSS are outlined as follows:

- Creation and identification of business agglomerations from existing business firms within a geographical region.
- Construction of intraspecific competition map between business firms in each business agglomerations.
- Predicting performance of a new business firm in each business agglomerations.
- Selecting the most appropriate business agglomeration for the new business firm based on its predicted performance.

However, in this study some modifications are made to improve the quality of the decision being made. These modifications are applied in the business agglomerations identification phase and in the final stage of selecting the final business location for a new business firm.

In the business agglomerations identification phase a graph-clustering algorithm is put into place to create clusters of business firms. In this study, a graph-clustering algorithm that partitions vertices by taking into account not only distances between vertices but also linkage between vertices, as proposed by [9] is used as an alternative to the *r*NN technique used by [6]. It is expected that considering the existence of linkage between vertices helps to construct more consistent groups of business firms.

Additional adjustment is also made to the rule of selecting the best business agglomeration. Previously selection was made based only on the prediction of the new business firm performance that is its predicted level of revenue, when being placed in a particular agglomeration. Conversely, this study considers conditions of existing business firms in an agglomeration after a new establishment joins that particular group to make the final decision of selecting a new business location. When a new business firm joins a business agglomeration then it is logical to consider that the existence of this new business firm may have effects to the performance of existing business firms. The best selection of business location should be the agglomeration in which a new business firm performs considerably well whilst at the same time performance of existing business firms are maintained.

The graph-clustering process applied in this study to construct business agglomerations from existing business firms in a geographical area is outlined as follows:

- Step 1, let a set of business firms is denoted by a graph $G(V,E)$ where V represents n numbers of business firms in an area whilst E represents geographical distances between these business firms. An n x n adjacency matrix A of a fully connected-graph then can be developed to represent this graph. Different types of distance measurement, i.e. Euclidean distance, Haversine formula, City Block distance, etc., can be used to calculate distance between business firms.
- Step 2, the initial process of constructing the business agglomerations is to remove links between distant business firms. It is assumed here that distant business firms are more likely to be placed in different agglomerations. Additionally, competitions between distant business firms are also being considered to be weak enough that they can be ignored. Maximum distance, *mdist* that allows a connection between two business firms to be maintained is defined by the lower quartile value of business firm distances distribution. A new adjacency matrix *Anew* can then be constructed by performing a matrix sparsification process. If $A_{ij} > mdist$ then $Anew_{ij} = \infty$, if $A_{ij} \leq mdist$ then $Anew_{ij} = A_{ij}$ and $Anew_{ij} = 0$ if $i = j$.

- Putting a business firm with its nearest neighbor to a same agglomeration then assembles Step 3 of the process, construction of initial clusters. For $i = 1$ to n, find nearest neighbor of business firm i. If business firm i does not belong to any agglomeration then join i and its nearest neighbor into one cluster, otherwise put nearest neighbor of business firm i in the same cluster as business firm i. This step creates the initial business agglomerations.
- Step 4, merging the initial clusters to form s business agglomerations by taking into account linkage or the intraspecific competition factor between business firms.
  Given that a quality vector Q consisting of m quality factors $(q_1,q_2,...,q_m)$ for a set of business firms operating in similar line of business is known, an affinity matrix W can then be constructed as follows:

$$\boldsymbol{W_{ij}} = \boldsymbol{P_{ij}}\exp\left(-\frac{\|\boldsymbol{q_i} \;-\; \boldsymbol{q_j}\|^2}{2\boldsymbol{\sigma}^2}\right), \qquad (1)$$

  where Pij = 1 if $Anew_{ij} \neq 0$ and $P_{ij} = 0$ if $Anew_{ij} = 0$. q_i and q_j represents quality vector of business firm i and j whilst σ is the standard deviation of quality vector of existing business firms. Higher value of w_{ij} indicates stronger competition or linkage between two establishments.
  The merging process of the initial clusters is then performed by evaluating the connectivity in terms of competition index between members of two different agglomerations. If significant connectivity is identified between members of two different clusters then a new business agglomeration is formed by merging members of those clusters. The merging procedure applied in the developed DSS is the cluster-merging algorithm, which is based on the maximum graph structural affinity of two groups as proposed by [9].
  A number of final business agglomerations s is then created in the completion of the process.

The set of business agglomerations created in the graph-clustering process is the candidate in which a new establishment can be put. As it has been explained in previous section, the term location in this study does not refer to a specific address of location neither some geographical coordinates, yet it refers to a particular business agglomeration.

Next process of the developed DSS is to predict the performance of a new establishment when it joins a business agglomeration given the quality vector of this new business firm is known.As it was proposed by [6] the task of predicting performance of a new establishment can be converted to a classification task. A key requirement of converting such task to become a classification task is to translate performance of each business firms into a discrete value or distinct categories as a replacement for the use of continuous number, i.e. amount of sales, revenue, profit, and so forth.

By transforming the performance indicator score of a business firm to a discrete value or distinct category then the task of predicting performance of a new establishment in an agglomeration is transformed to an estimating performance category task where the members of the business agglomeration whose the quality vector and performance category are known serve as the training set. In this case the predicted class label will act as the projected performance category of the new establishment in a particular agglomeration.

Detail procedures of the algorithm implemented in the developed DSS is described as follows:

- Step 1, after the graph-clustering process creates s business agglomerations, then the first step to be performed in predicting performance category of a new business firm in each agglomerations is to construct the extended affinity matrix *Wext* with the size of $(n+s)$ x $(n+s)$. The key idea of this extended affinity matrix construction is to represent the state where a new establishment e is being put in each business agglomerations.
  Consequently affinity indices between new establishments (we consider there are s new establishments as they are s business agglomerations) with other members of a business agglomeration need to be calculated given the quality vector of the new business firm is known.
  The extended affinity matrix is constructed as in (1) for $i = 1$ to n and $j = 1$ to n, whilst for i = 1 to n and $j = (n+1)$ to $(n+s)$ $Wext_{ij}$ is only being calculated if business firm i is a member of agglomeration $(n+1),(n+2),\ldots,(n+s)$. For $i = (n+1)$ to $(n+s)$ and $j = 1$ to n, $Wext_{ij}$ is also only being calculated if business firm j is a member of agglomeration $(n+1),(n+2),\ldots,(n+s)$.
- Step 2, once the extended affinity matrix has been constructed then the next step is to calculate the projection of performance category for e in each particular agglomeration. As proposed by [6] in their work, the projected performance category can be calculated as follows:

$$\boldsymbol{F}^* = (\boldsymbol{I} - \boldsymbol{\alpha S})^{-1}\boldsymbol{Y}. \qquad (2)$$

  Y is the performance category matrix with element pc for each establishment with the size of $(n+s)$ x l where l is the number of distinct performance category.

$$\boldsymbol{Y_{ij}}\begin{cases}\boldsymbol{1}, \boldsymbol{pc_i} = \boldsymbol{j} \\ \boldsymbol{0}\end{cases}. \qquad (3)$$

  α is a multiplier factor defined by the following equation:

$$\boldsymbol{\alpha} = \frac{\boldsymbol{1}}{\boldsymbol{1}+\boldsymbol{\mu}}, \qquad (4)$$

  where μ is the mean value of S. Consequently, S is the normalized *Wext* computed as follows:

$$\boldsymbol{S} = \sqrt{\boldsymbol{D}} \times \boldsymbol{Wext} \times \sqrt{\boldsymbol{D}} \quad . \qquad (5)$$

D is a diagonal matrix where D_{ij} is calculated as follows:

$$D_{ij} = \sum_{i=1}^{m} Wext_i, \quad (6)$$

for $i = j$ and m is the size of $Wext$.

The F^* matrix represents the projected performance category for the new business firm in each business cluster and projected performance category for the other members of the agglomeration when the new establishment joins them.

Dimension of F^* is $(n+s)$ x s and the selection of the most appropriate agglomeration for the new establishment would be agglomeration j, where j is the column of F^* in which the projected performance category of the new establishment scores maximum. A projected performance category matrix $Ynew$ can then be calculated where for each row of $Ynew$ column j is 1 if F^*j is maximum for that i-th row and 0 otherwise.

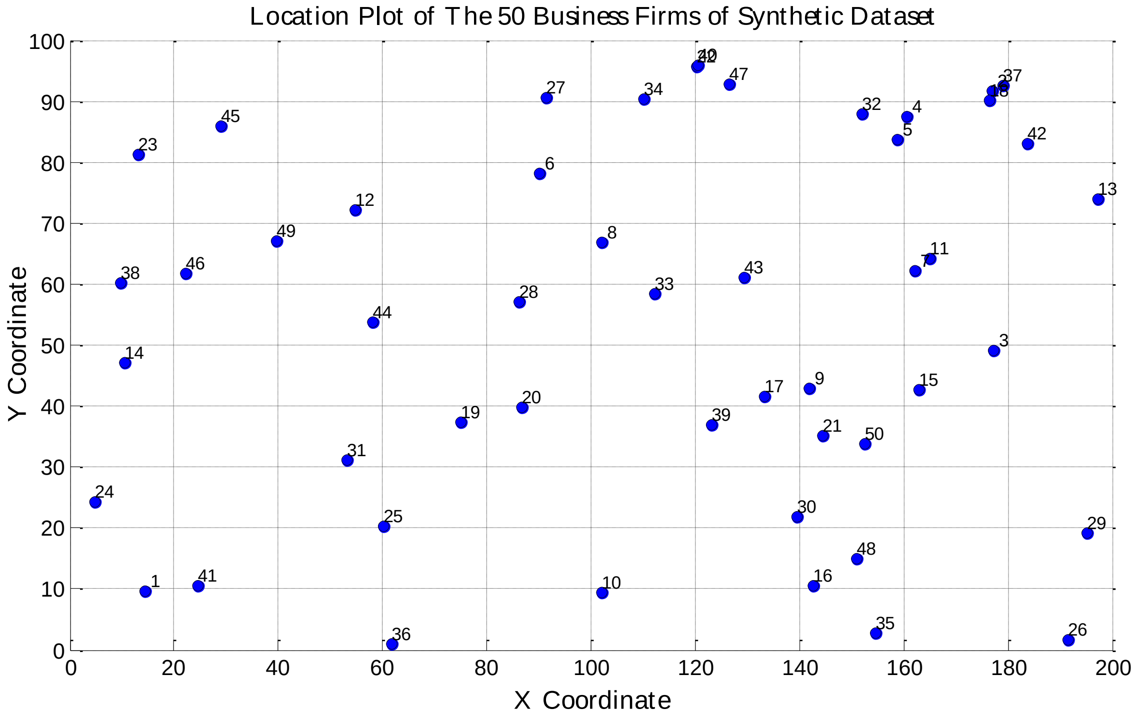

Figure 1. Distribution of the 50 business firm in space domain where x ranges from 0 to 200 and y ranges from 0 to 100.

TABLE I. ATTRIBUTES OF THE 50 BUSINESS FIRMS OF THE SYNTHETIC DATASET

Firms ID	Q1	Q2	Category	Firms ID	Q1	Q2	Category	Firms ID	Q1	Q2	Category
1	2.50	1.01	4	**18**	2.49	1.77	2	**35**	1.51	2.93	3
2	4.70	2.26	3	**19**	3.10	1.40	3	**36**	2.41	4.37	5
3	3.81	2.52	3	**20**	4.63	4.05	3	**37**	2.83	4.16	2
4	3.22	4.06	2	**21**	3.97	3.16	4	**38**	3.73	1.25	5
5	4.13	1.55	3	**22**	1.35	2.00	4	**39**	2.85	3.52	4
6	4.50	2.35	1	**23**	3.53	2.33	3	**40**	2.01	1.97	4
7	1.06	1.50	4	**24**	1.63	3.26	1	**41**	1.07	1.09	4
8	1.18	4.91	5	**25**	4.40	2.85	4	**42**	4.75	1.92	1
9	3.40	2.10	1	**26**	1.95	1.27	2	**43**	2.82	1.35	2
10	2.30	4.29	1	**27**	3.42	2.16	4	**44**	2.88	2.07	4
11	2.25	4.28	5	**28**	3.12	4.25	5	**45**	4.20	4.24	3
12	3.25	1.26	2	**29**	1.66	3.26	1	**46**	1.85	4.78	5
13	1.40	4.07	1	**30**	1.47	4.67	5	**47**	4.17	4.48	4
14	4.82	1.70	5	**31**	1.06	1.58	1	**48**	1.16	2.11	3
15	1.03	1.79	1	**32**	4.88	4.77	5	**49**	2.77	4.03	2
16	4.41	4.02	5	**33**	4.76	1.46	5	**50**	3.24	3.49	2
17	3.44	2.72	1	**34**	4.17	1.75	1				

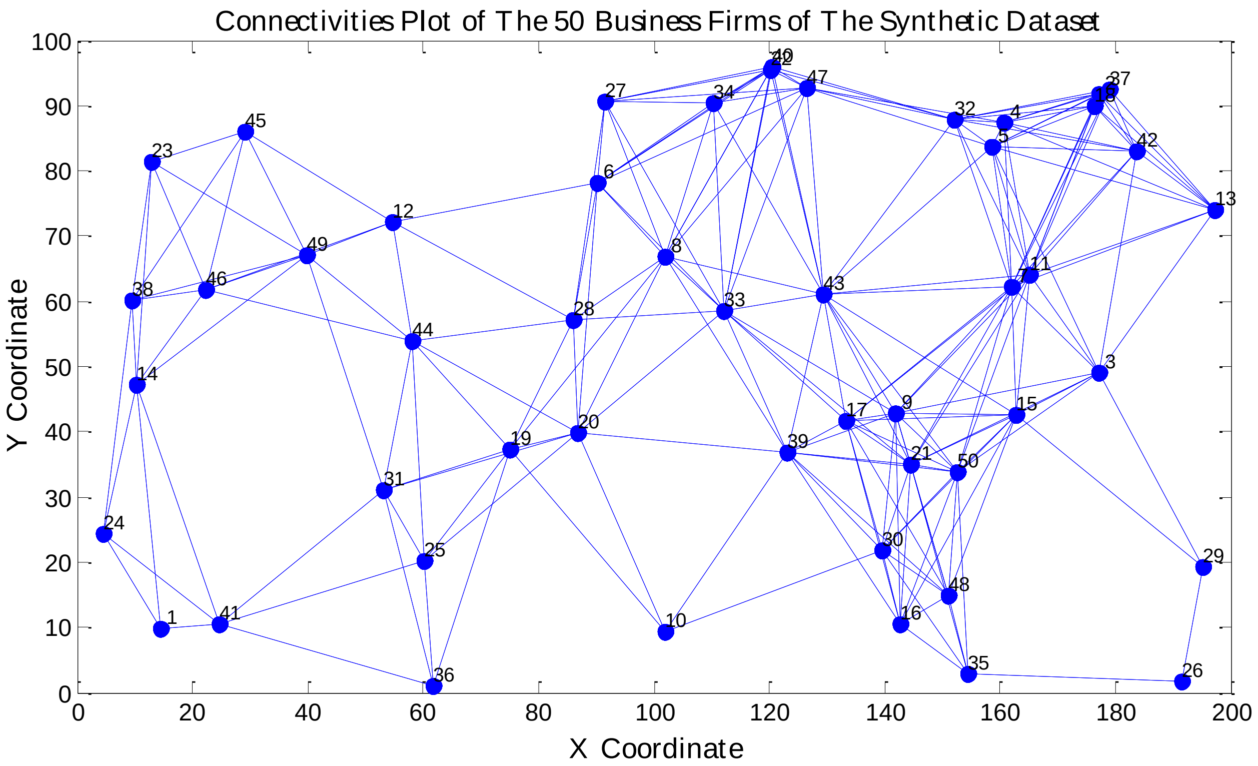

Figure 2. Connectivity graph between the 50 business firms of the synthetic dataset adter sparsification process.

In addition, considering the projected performance category of existing business firms also forms the final selection of a business agglomeration. An agglomeration is selected only if after the propagation process no significant decrement in the performance category of existing business firms is detected. This analysis can be done by comparing the value between the *Y* matrix and the *Ynew* matrix for $i,j = 1$ to n.

Both the graph-clustering process and the projection of performance category described in this section are the core methods of the developed DSS to identify the most appropriate business agglomeration for a new business firm to be established.

Upcoming section of the paper outlines result of conducted experiment of the developed DSS using synthetic dataset that mimics a real condition of business units' distribution in a geographical area.

IV. EXPERIMENTAL STUDY AND DISCUSSION

A synthetic dataset is constructed to help validating the performance of developed DSS in this study. This synthetic dataset consists of 50 business units, which are spread out in a geographical region randomly. The location of each business firm is represented by coordinates (x,y) with a range of value between 0 to 200 for x and 0 to 100 for y. Additionally, a quality vector q formed by two quality components q_1 and q_2 are given for each establishment. In the constructed dataset both q_1 and q_2 ranges from 1.00 to 5.00 for each establishment, where greater number indicates better quality.

Furthermore, a performance label or each business firm in the synthetic dataset is also defined. Figure 1 shows the spreading of the 50 business units in a space domain whilst detail attributes of the dataset are outlined in Table I.

It is expected that the developed DSS is capable of identifying trustworthy business agglomerations and then to calculate the projected performance category for the new establishment and existing business firms as well.

Figure 2 illustrates the connectivites between existing businesses units after insignificant linkages are removed based on distance threshold. It clearly shows that the algorithm maintain only connection between nodes which are considerably close. Figure 3 describes the initial agglomerations constructed from the existing 50 business firms. As described in previous section, these initial clusters are developed simply by merging business units with their nearest neighbor in terms of geographical distance. It can be observed that members of each agglomeration in these initial clusters are those, which are physically, can be considered as neighboring business units.

Final business agglomerations constructed by the cluster-merging algorithm from the 50 business units of the synthetic dataset is shown in Figure 4. As a result only 5 clusters instead of the initial 15 clusters are recognized after the cluster-merging process.

It can be analyzed from Figure 4 that the algorithm is capable to merge initial clusters of business units to form reliable conclusive business agglomerations by taking into account the affinity indices between initial agglomerations.

Given the quality vector of a new establishments, q_{new} whereas $q_{new1} = 1$ and $q_{new2} = 3$, projected performance category of this new establishment is then calculated for each agglomerations.

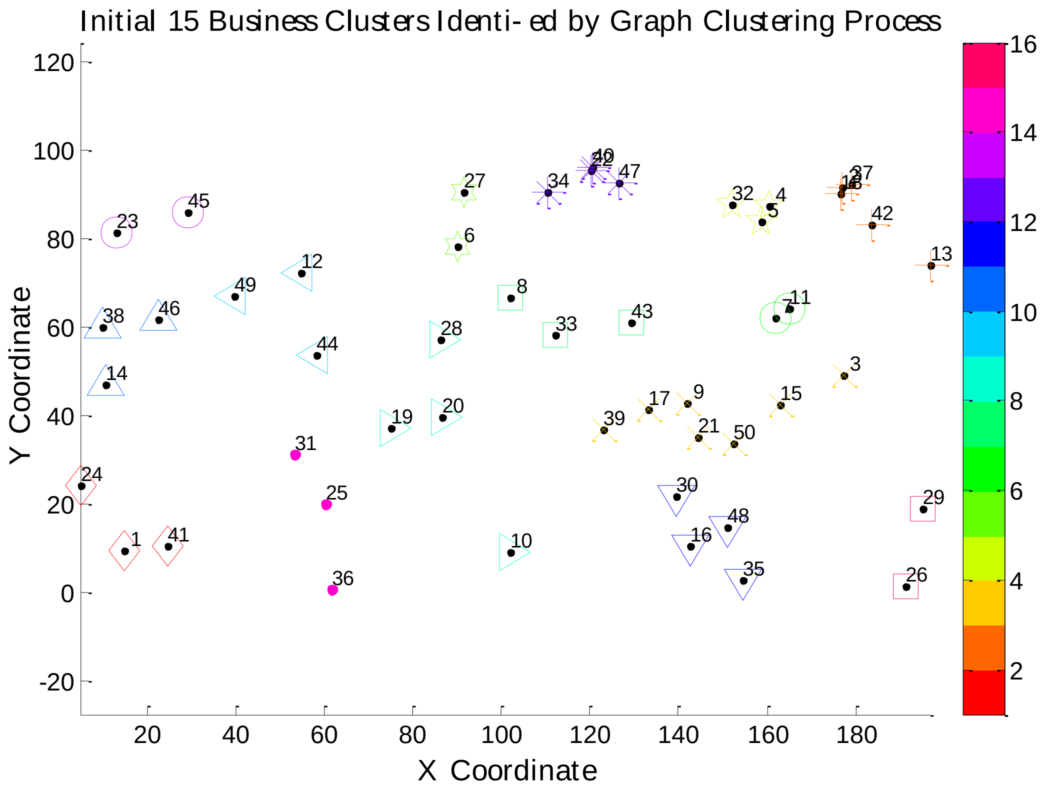

Figure 3. Initial 15 business agglomerations identified by the graph-clustering process for the 50 business firms in the synthetic dataset before cluster-merging process is applied.

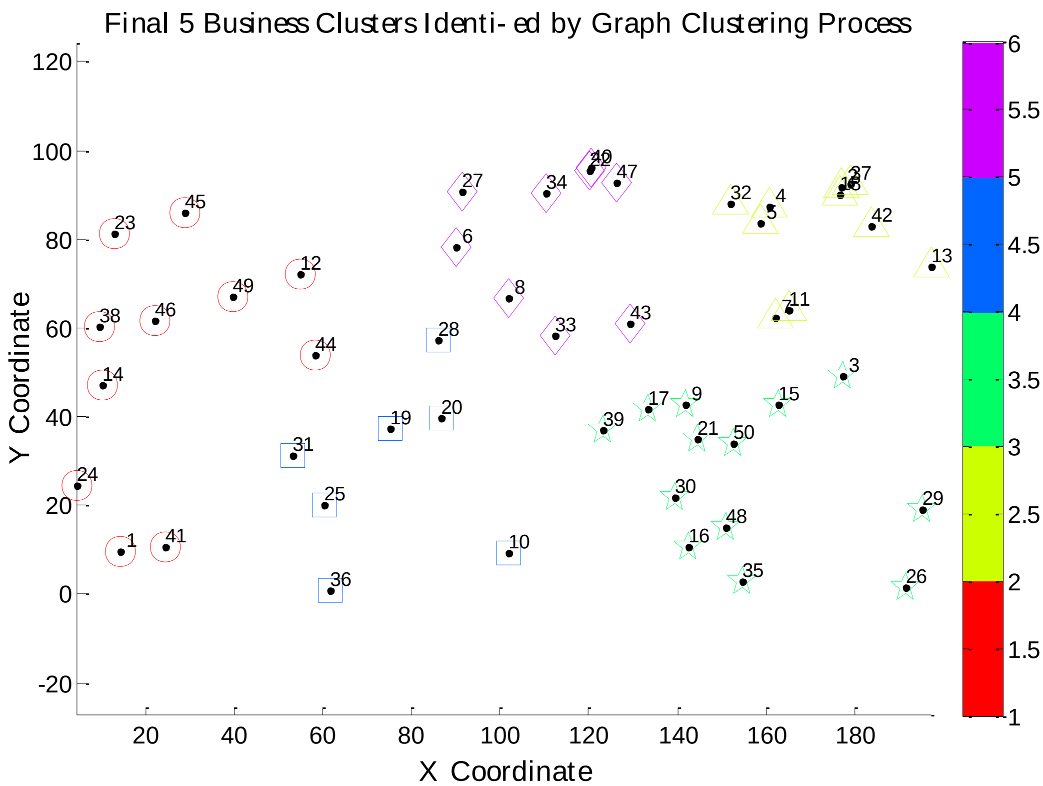

Figure 4. Final 5 business agglomerations identified by the graph-clustering process for the 50 business firms in the synthetic dataset.

TABLE II. PROJECTED PERFORMANCE LABEL OF NEW ESTABLISHMENT IN EACH IDENTIFIED BUSINESS AGGLOMERATIONS

BA #1	BA #2	BA #3	BA #4	BA #5
4	1	1	1	4

Based on results outlined in Table II, the developed DSS suggests that the new business firm should join Business Agglomeration #1 or Business Agglomeration #5 in which it scores the highest performance label. However, further analysis related to the projected performance category of existing business firms the developed DSS proposes that the new establishment should join Business Agglomeration #1 since when it joins Business Agglomeration #5 there is a significant decrement in the performance category of members of Agglomeration-5. Comparing the Y matrix, which represents the initial performance category for the whole 50 business units and the *Ynew* matrix that indicates the condition of these business firms after the new establishment joins them serves as the basis of this analysis.

V. CONCLUSION AND FUTURE WORK

Results of conducted experiment using synthetic dataset in this study indicate that the developed DSS could be of help in selecting a business location for a new establishment.

A graph-clustering algorithm that considers the affinity indices between nodes or business firms in this study is proven to be able to construct trustworthy business agglomerations by merging the initially formed clusters. Additionally, taking into account the projected performance category of each existing business firms when selecting an agglomeration for a new establishment helps to confirm that the existence of the new establishment will not put existing business firms in danger. This is important since the economy of a particular region will grow better if the formation of new business firms helps the existing business firms to develop as well.

This study is an initial step to build an applied DSS that is capable of helping the Indonesian Government in planning the development of business units in a geographical area. It is expected that by being able to do so would help to escalate the economy growth in Indonesia. Consequently, the future work defined for this study is to apply the developed DSS to real-world business firms' data, which in this case the small and medium enterprises data collected from other regions in Indonesia. Furthermore, a user-friendly application is to be developed so the system can be easily used and maintained by the Indonesian Government.

ACKNOWLEDGMENT

This study is funded by the Directorate of Higher Education, Education and Culture Ministry of the Republic of Indonesia through the Strategis Nasional Research Grant No. 159/SP2H/PL/DIT.LITABMAS/V/2013

REFERENCES

[1] K. Marijan, "Mengembangkan Industri Kecil Menengah Melalui Pendekatan Kluster" in INSAN, 7(3), 2005, pp. 216–225.

[2] M. Fujita, M. and J.F. Thisse, "Economics of Agglomeration," in J. Japanese and International Economics, December 1996, pp. 339–378.

[3] A. Marshall, Principles of Economics. Macmillan, 1980.

[4] F.S. Steingold, Legal Guide for Starting & Running a Small Business Vol. 1. Nolo Press, 1999.

[5] R. Hayter, The Dynamics of Industrial Location: The Factory, the Firm, and the Production System. John Wiley & Sons, 1997.

[6] X. Quan, L. Wenyin, W. Dou, H. Xiong and Y. Ge, "Link Graph Analysis for Business Site Selection," in IEEE Computer Society's Computer, 0018-9162, 2012, pp. 64–70.

[7] Y. Zhou, H. Cheng and J. X. Yu, "Graph Clustering Based on Structural/Attribute Similarities," in Proceedings of VLDB '09, August 2009, Lyon, France.

[8] T. Joachims, "Transductive learning via spectral graph partitioning, " in Proceedings of the International Conference on Machine Learning, 2003, pp. 290–297.

[9] W. Zhang, D. Zhao and X. Wang, "Agglomerative clustering via maximum incremental path integral," in IEEE Pattern Recognition, 46 (11), 2013, pp. 3056–3065.

[10] A. Blum and T. Mitchell, " Combining labeled and unlabeled data with co-training," in Computational Learning Theory, 1998, pp. 92–100.

[11] X. Zhu, Semi-supervised learning with graphs. Technical Report, Carnegie Mellon University, Pittsburgh, PA, 2005.

[12] X. Zhu, Semi-Supervised Learning Literature Survey. Techical Report 1530, Dept. of Computer Sciences, University of Wisconsin-Mad, 2006.