“Optimized consortium formation through cluster analysis”

AUTHORS	Kgwadi M. Mampana Solly M. Seeletse Enoch M. Sithole
ARTICLE INFO	Kgwadi M. Mampana, Solly M. Seeletse and Enoch M. Sithole (2016). Optimized consortium formation through cluster analysis. *Problems and Perspectives in Management*, *14*(1), 117-126. doi:10.21511/ppm.14(1).2016.13
DOI	http://dx.doi.org/10.21511/ppm.14(1).2016.13
RELEASED ON	Wednesday, 02 March 2016
JOURNAL	"Problems and Perspectives in Management"
FOUNDER	LLC “Consulting Publishing Company “Business Perspectives”

Kgwadi M. Mampana (South Africa), Solly M. Seeletse (South Africa), Enoch M. Sithole (South Africa)

Optimized consortium formation through cluster analysis

Abstract

Some problems cannot be solved optimally and compromises become necessary. In some cases obtaining an optimal solution may require combining algorithms and iterations. This often occurs when the problem is complex and a single procedure does not reach optimality. This paper shows a conglomerate of algorithms iterated in tasks to form an optimal consortium using cluster analysis. Hierarchical methods and distance measures lead the process. Few companies are desirable in optimal consortium formation. However, this study shows that optimization cannot be predetermined based on a specific fixed number of companies. The experiential exercise forms an optimal consortium of four companies from six shortlisted competitors.

Keywords: distance measures, hierarchical methods, optimal consortium.
JEL Classification: C1, C3, C4, C5, C6.

Introduction

Combinations of entities for working together are sometimes inevitable. The aim of combining is to create synergies for improved performance. These combined entities do not always result in the outcome desired. Hence, it is vital that when such combinations are formed, mechanisms should be designed to enhance that they perform at the required levels. Among the common and also important combinations that have shown failures in recent times are the public-private partnerships and consortia (Joshi, 2010). A consortium is a conglomerate of several entities working together towards a collective objective (Dolnicar & Lazarevski, 2009). Companies form a consortium through the process of clustering almost daily. Some consortia succeed while others fail (Larson et al., 2005). Partnerships fail because of '*lack of chemistry*' between the component entities (Koti, 2006). This study designs an attraction in consortium formation. Using systematic methods enhances consortium success, and not using them heightens chances for consortium failure. This paper applies cluster analysis to form optimal consortia.

1. Cluster analysis techniques

The purpose of cluster analysis is to discover a system of organizing entities into groups in which group members share properties (Seo & Shneiderman, 2002). This study applies cluster analysis to determine optimal consortia. Involved companies should form synergies. In the sense of this paper, attributes in the optimal consortium should possess the best possible performance promise.

Kgwadi M. Mampana, Ph.D., Department of Statistics, University of South Africa, Florida Campus, South Africa.
Solly M. Seeletse, Ph.D., Professor, Department of Statistics and Operations Research, Sefako Makgatho Health Sciences University, South Africa.
Enoch M. Sithole, Ph.D., Department of Physics, Sefako Makgatho Health Sciences University, South Africa.

1.1. Fundamental cluster analysis steps. Cluster analysis starts from a proximities matrix of the items to be grouped (Everitt, Landau & Leese, 2001). It combines items such that grouped items do not include duplications, rather, items should complement one another (Kaufman & Rousseeuw, 1990). Grouped items should add value by enabling synergies, and strengths of one item should offset weaknesses of others. Cluster analysis starts with a data matrix displaying the column of items and rows depicting criteria (or attributes) that is converted into a proximities matrix (Dhillon & Modha, 2001). The proximities matrix shows proximity values of the different items while its diagonal elements are all zero to indicate no distance between an item and itself. Two ultimate tasks are imperative. Firstly, a decision is required about the items to be gathered for inclusion. Secondly, the method to apply in combining multiple measures into a single similarity measure between the items should be decided. A typical data matrix takes the normal known form of rows and columns as follows:

Table 1. Data matrix format

	X1	.	.	.	An
A1	x11				x1n
.	.				.
.	.				.
.	.				.
Am	x1m				xnm

1.2. Hierarchical methods and distances in grouping items. Similarity linkages and distances measures are used to compare items to be clustered. Derived measures apply in cluster analysis by grouping the items into clusters. Few clusters are desirable, but they should be adequate to possess a desirable number of attributes for the tasks required. If the number of items for a cluster/consortium is not known beforehand, hierarchical linkage methods are useful (Kraskov et al., 2003). The linkage methods are discussed next, followed by distances.

Single linkage: Single linkage is based on clustering items that are nearest neighbors. It calculates the distance between two items as the minimum distance between any two items. Allocation is done for the first pair which shows the minimum distance, and the scores are combined as applicable. Then the distances for the new groups are calculated again. The next allocation is based on the shortest distance that emerges at this stage. The process is repeated for the subsequent steps. Clustering is considered complete when an optimal state is achieved.

Complete linkage: Complete linkage is the opposite in use of distance to the single linkage as it uses the furthest neighbor as the criterion for clustering items. It also starts by calculating the distances between pairs of items in each step. It then groups together items that show to have the maximum distance.

Average linkage: The procedure of average linkage is similar to the single and maximum linkage methods, but considers the average distance. It computes the distance between subgroups at each step as the average of the distances between the two items. The procedure continues as for the previous linkages methods, and the process stops when optimality is achieved.

Mahalanobis distance

The Mahalanobis distance (MD) measures the distance between two correlated variables (Weisstein, 2003). Let $N_p(\mu, \Sigma)$ be the probability density function of the normal distribution. The MD measure between x and y in the p-dimensional space is given by:

$$\delta^2(x,y) = (x-y)'\Sigma^{-1}(x-y). \tag{1}$$

Geometric distance

Geometric distances are often measured in the Euclidean space, where a distance is a numerical description of the way items are lying far apart (Seker, Altun, Ayan & Mert, 2014). Distance is a metric function to describe that items are “close to” or “far away from” each other. In real numbers, metric distance between x and y satisfies the conditions:

- $d(x,y) \geq 0$,
- $d(x,y) = 0 \Leftrightarrow x = y$;
- $d(x,y) = d(y,x)$;
- $d(x,z) \leq d(x,y) + d(y,z)$.

Analytic geometry definition of the Euclidean distance between two items is:

$$d = \sqrt{\sum_{i=1}^{p} \left(x_{ji} - x_{ki}\right)^2}. \tag{2}$$

Manhattan distance

When p = 1, the following formula results are the Manhattan distance:

$$d_1 = \sum_{i=1}^{p} \left|x_{ji} - x_{ki}\right|. \tag{3}$$

Euclidean distance

The case p = 2 yields the Pythagorean Theorem generalization:

$$d_2 = \left(\sum_{i=1}^{p} \left|x_{ji} - x_{ki}\right|^2\right)^{\frac{1}{2}}. \tag{4}$$

The *p*-norm

$$d_p = \left(\sum_{i=1}^{p} \left|x_{ji} - x_{ki}\right|^p\right)^{\frac{1}{p}}. \tag{5}$$

Chebyshev distance

The Chebyshev distance is defined by:

$$d_\infty = \lim_{p\to\infty} \left(\sum_{i=1}^{p} \left|x_{ji} - x_{ki}\right|^p\right)^{\frac{1}{p}}. \tag{6}$$

Matthews correlation coefficient

The machine learning description of the Matthews correlation coefficient (MCC) is that MCC is a measure of the quality of binary (two-class) classifications (Perruchet & Peereman, 2004; Powers, 2011). The MCC allows for true and false positives and negatives. Fawcelt (2006) describes the MCC as a correlation coefficient between the observed and predicted binary classifications ranging from −1 to +1. Let n be the total number of observations. The MCC statistic (or the phi-coefficient) is:

$$MCC = \sqrt{\frac{\chi^2}{n}}. \tag{7}$$

1.3. Fuzzy logic techniques.

Decision making under pure uncertainty

Personality type and decision making work together. People often make decisions due to their inner influences (Triantaphyllou, 2000). When a person controls a system fully, and is influential, their approach tends to depend on their ‘basic’ expectations. Focus is on ‘pessimists’ and ‘optimists’ to complement the methods presented earlier. ‘Pessimism’ expects that bad things always happen, and considers the possible worst cases of all the alternatives. It starts by selecting the alternative with the minimum payoff, and then selects the maximum of the minima (MaxMin process, or maximizing the minimum possible gain). ‘Optimism’ is the approach of maximum of the maximum gains (MaxMax process, or maximizing the maximum possible benefit).

Regret approach

The approach in decision making is to minimize risks (Sharma, 2006). Thus, reducing regrets becomes important. Regret is the payoff on what would have been the best decision in the circumstances minus the payoff for the actual decision in the circumstances. Therefore, the first step is to setup the regret table:

- Take the largest number in each states of nature column.
- Subtract all the numbers in that state of nature column from it.
- Choose maximum number of each action.
- Choose minimum number from previous step and take that action.

Expectations

The expected payoff (EP) approach requires estimating the expectation and then selecting the maximum expected pay-off. The expected opportunity loss (EOL) is the expected loss of an opportunity that would have yielded a greater benefit. Risk assessment is a procedure of quantifying the loss or gain values and supplying them with proper probability values (Bergman, 2009). Smaller values of risk indicate that what is expected is likely to be what occurs. The states of nature are the states of economy during an arbitrary time frame. The expected value needs conditions for good indication of a quality decision. The variance is an important measure of risk. A large expected return is desired, with small risk. Large variance indicates a higher risk for the system being measured. The CV is a useful relative risk (Limpert, Stahel & Abbt, 2001) based on mean and standard deviation expressed as:

$$CV = \frac{standard \quad deviation}{mean}. \tag{8}$$

A smaller CV indicates more reliability. Thus, data with a smaller CV are more stable (i.e. lower risk).

2. Data

In an exercise in which a tender invitation was issued, several companies were evaluated using scores on nine project attributes to determine the winner. Each attribute was judged out of 100. No competitor was found adequate, but some had attributes indicating promising performance on some aspects of the identified project. Also, when combined, they contained all the desirables. The scores of the top six shortlisted companies were used to form consortia in identifying the optimal consortium.

Table 2. Data matrix

	A1	A2	A3	A4	A5	A6	A7	A8	A9
C1	57	11	23	63	64	51	60	15	62
C2	43	51	14	29	57	48	35	31	28
C3	39	60	61	41	43	33	36	42	34
C4	65	38	35	53	19	67	53	15	57
C5	17	26	41	26	21	22	15	55	44
C6	22	49	52	21	33	18	15	64	16

3. Findings

3.1. Company mean strengths. The averages (also expected pay-offs) of the points awarded to each company are considered at this stage to determine the rated performances of these companies.

Table 3. ANOVA: 1-factor without replication

Entity	Sum	Average	Variance	CV
C1	406	45.1111	489.8611	0.4906
C2	336	37.3333	183.2500	0.3626
C3	389	43.2222	107.9444	0.2404
C4	402	44.6667	360.0000	0.4248
C5	267	29.6667	189.0000	0.4634
C6	290	32.2222	334.4444	0.5676

The 4^{th} column of Table 3 shows average scores of company strengths. Merit order is C1 (45.11 points); C4 (44.67 points); C3; C2; C6; and C5. Leader C1 has most attributes in which it leads all the others, but performs poorly at attributes A2, A3 and A8. It is the winning candidate, but should be clustered to offset its weak parts. At C1’s weakest attributes, C3 leads at A2 and A3; C6 leads at A8, and C4 leads at A1. C2 and C5 are not leading at any attribute, and cannot improve the weaknesses of other companies. They are candidates for exclusion in any consortium. Initial possible cluster pairs could be (C1:C3), (C1:C6) and (C4:C6).

3.2. Consideration of relative company stability. CVs measure the stabilities of the companies (Table 3). Low performer C6 and high performer C1 have highest CVs, indicating highest instabilities. C3 has the least CV (is most stable). Thus, C3 is most trusted with least risk. C3 is also a high performer. Hence, these count for C3 inclusion. Poor performer C2 has the next smallest CV value, but is not considered because of its poor performance. Company C4 is the next highest stable company, based on CV. Leading unstable companies are C6, C5 and C1. Like C2, poor performer C5 is excluded. The weak attributes in high performer C1 are points of weakness requiring to be strengthened. Low performer C6 performs extremely high at attribute A8. This could offset the A8 weakness of partner companies when included in a cluster.

Statistical comparison of mean strength of companies

Table 4. ANOVA

Variation source	*SS*	*df*	*MS*	*F*	*p*-value	*F* crit
Rows	2003.259	5	400.6519	1.2148567	0.319919	2.449466
Columns	124.2593	8	15.53241	0.0470974	0.999942	2.18017
Error	13191.74	40	329.7935			

Interpretation of the ANOVA table

The null hypothesis of ANOVA test is that the means are all equal. The ANOVA table (Table 4) indicates that the values of the points awarded to the companies are not significantly different, based on the p-value exceeding 0.05. The table also does not indicate the differences in the strengths obtained on the attributes during shortlisting. This information therefore, shows that the mean strengths of the companies are not significantly different.

3.3. Statistical comparison of mean strength of companies.

Table 5. Correlation matrix of companies

	C1	C3	C4	C6
C1	1			
C3	-0.7318	1		
C4	0.5073	-0.4745	1	
C6	-0.8920	0.7115	-0.8027	1

Multicollinearity is shown for pairs (C1:C3), (C1:C6), (C3:C6), (C4:C6). C3 and C6 have a high positive correlation indicating that the two companies are similar and cannot add value to each other when merged together. Consideration for clustering point at (C1:C3), (C1:C6) and (C4:C6).

Table 6. Proximities matrix for companies

	C1	C3	C4	C6
C1	0			
C3	1.89	0		
C4	0.44	1.45	0	
C6	12.89	11.00	12.45	0

The pairs of companies based on smallest distances are (C1:C4), (C3:C4), (C1:C3), (C3:C6), (C4:C6), (C1:C6). Pair (C1:C4) is undesirable at A2 performance of at most 38, A3 of at most 35, and A8 of at most 15. Another pair in which optimality can similarly not be reached is (C3:C6) because they can only perform at A2, A3, A8. Promising pairs towards optimal consortium are: (C3:C4), (C1:C3), (C4:C6), (C1:C6).

3.4. Applying distance measures. The p-norm distance for the current problem is:

$$d = \left(\sum_{i=1}^{9} \left| x_{ji} - x_{ki} \right|^{9} \right)^{\frac{1}{9}}$$

Applying this on the data matrix the proximities matrix is:

Table 7. Proximities matrix

	C1	C3	C4	C6
C1	0			
C3	49.60	0		
C4	45.05	35.29	0	
C6	54.21	24.22	54.56	0

Single linkage: Minimum distance occurs between C3 and C6. Thus, the next proximities matrix starts by merging C3 with C6. The new distance is the minimum distance between any company with C3 or C6:

Table 8. Proximities matrix

	C1	C3:C6	C4
C1	0		
C3:C6	49.60	0	
C4	45.05	35.29	0

The next cluster is C3:C4:C6 based on minimum distance. Then the next proximity matrix is:

Table 9. Proximities matrix

	C1	C3:C4:C6
C1	0	
C3:C4:C6	45.05	0

The two items resulting are C1 and C3:C4:C6. Even though it is not material at this stage, the distance between these consortia is 36.91, which is the minimum (45.05; 36.91) of the two distances involved. The new data matrix of strengths evolving from Table 3 becomes:

Table 10. Single linkage-based data matrix

	A1	A2	A3	A4	A5	A6	A7	A8	A9
C1	57	11	23	63	64	51	60	15	62
C3:C4:C6	65	60	61	53	43	67	53	64	57

These two best items do not define optimal consortia. The first one is outwitted at attributes A1 to A3, A6 and A8. Elsewhere in other attribute the second one is outwitted. The aim is to find optimal solution to consortium formation with the smallest possible number of companies. Since at this stage optimality has not materialized, the process continues.

Complete linkage: From Table 7, closest companies based on closest proximities are C3 and C6. The next proximities matrix is obtained by merging C3 with C6, forming cluster C3:C6. The new distance between other members and the cluster is the maximum distance between any company with C3 and C6 as follows:

Table 11. Proximities matrix

	C1	C3:C6	C4
C1	0		
C3:C6	54.21	0	
C4	45.05	54.56	0

Members close to each other from Table 11 are C1 and C4. The next proximities matrix has cluster C1:C4 and C3:C6. The new distance from the cluster to any other member is the maximum distances as:

Table 12. Proximities matrix

	C1:C4	C3:C6
C1:C4	0	
C3:C6	54.21	0

The new data matrix of strengths becomes:

Table 13. Complete linkage-based data matrix

	A1	A2	A3	A4	A5	A6	A7	A8	A9
C1:C4	65	38	35	63	64	67	60	15	62
C3:C6	39	60	61	41	43	33	36	64	34

Again, none of these two consortia is optimal. The first one is outwitted at attributes A2, A3 and A8. Elsewhere in other attributes the second consortium is outwitted.

Average linkage: From Table 7, similar members are C3 and C6. Then the next proximities matrix is obtained by merging C3 with C6, forming cluster C3:C6. The new distance between other companies and the cluster is the average distance between any company with C3 or C6:

Table 14. Proximities matrix

	C1	C3:C6	C4
C1	0		
C3:C6	51.91	0	
C4	45.05	44.98	0

The smallest distance leads to the next new cluster C1:C3:C6. The results of the consortia formations according to the linkage methods become:

Table 15. Linkage-based consortia

	Single linkage	Complete linkage	Average linkage
Consortia	1. C1 2. C3:C4:C6	1. C1:C4 2. C3:C6	1. C1:C4 2. C3:C6

3.4.1. Euclidean distance. This distance measure is:

$$d = \sqrt{\sum_{i=1}^{p} \left(x_{ji} - x_{ki}\right)^2}$$

Table 16. Proximities matrix

	C1	C3	C4	C6
C1	0			
C3	86.6	0		
C4	58.2	72.5	0	
C6	117.8	49.6	107.0	0

Single linkage clustering leads to:

Table 17. Single linkage cluster matrix

	C1	C3:C6	C4
C1	0		
C3:C6	86.6	0	
C4	58.2	72.5	0

From the shortest distance, the next clusters are C1:C4 and C3:C6.

3.4.2. Manhattan distance. The Manhattan distance is:

$$d_1 = \sum_{i=1}^{p} \left|x_{ji} - x_{ki}\right|$$

Table 18. Proximities matrix

	C1	C3	C4	C6
C1	0			
C3	245	0		
C4	130	211	0	
C6	348	143	294	0

Complete linkage clustering leads to:

Table 19. Single linkage cluster matrix

	C1:C4	C3	C6
C1:C4	0		
C3	245	0	
C6	348	143	0

The next clusters, based on closest proximity, are C1:C4 and C3:C6.

3.4.3. The p-norm distance. The p-norm is

$$d_p = \left(\sum_{i=1}^{p} \left|x_{ji} - x_{ki}\right|^p \right)^{\frac{1}{p}}.$$

Table 20. Proximities matrix

	C1	C3	C4	C6
C1	0			
C3	49.6	0		
C4	45.1	35.3	0	
C6	54.2	24.2	54.6	0

Average linkage clustering leads to:

Table 21. Average linkage cluster matrix

	C1	C3:C6	C4
C1	0		
C3:C6	52.9	0	
C4	45.1	44.9	0

The next clusters, based on closest proximity, are C3:C6 and C1:C4.

3.4.4. Chebyshev distance. This measure is:

$$d_\infty = \max|x_{1i} - y_{1i}|, |x_{2i} - y_{2i}|, \ldots, |x_{ni} - x_{ni}|$$

Table 22. Proximities matrix

	C1	C3	C4	C6
C1	0			
C3	49	0		
C4	45	34	0	
C6	49	22	49	0

Single linkage clustering leads to:

Table 23. Single linkage cluster matrix

	C1	C3:C6	C4
C1	0		
C3:C6	49	0	
C4	45	34	0

The next cluster, based on closest proximity, is C3:C4:C6.

Matthews correlation

The Matthews correlation is:

$$MCC = \sqrt{\frac{\chi^2}{n}}.$$

Table 24. Proximities matrix

	C1	C3	C4	C6
C1	0			
C3	35.4	0		
C4	23.8	29.6	0	
C6	48.1	20.3	43.6	0

Complete linkage clustering leads to the next table.

Table 25. Single linkage cluster matrix

	C1	C3:C6	C4
C1	0		
C3:C6	48.1	0	
C4	23.8	43.6	0

The next clusters are C3:C6 and C1:C4. The consortia formed are as follows:

Table 26. Summary of consortia formation

Consortium	Frequency of occurrence	Number of times consortium came first
C1:C4	4	3
C3:C6	4	1
C3:C4:C6	1	0

The consortia revolve around C1, C3, C4, C6 and mostly as clusters C3:C6 and C1:C4.

Fuzzy logic techniques

MinMax approach

The approach requires identifying lowest performers in each attribute.

Table 27. MinMax matrix

	A1	A2	A3	A4	A5	A6	A7	A8	A9
C1	57	11	23	63	64	51	60	15	62
C3	39	60	61	41	43	33	36	42	34
C4	65	38	35	53	19	67	53	15	57
C6	22	49	52	21	33	18	15	64	16
Min	22	11	23	21	33	18	15	15	16
Company	C6	C1	C1	C6	C6	C6	C5:C6	C1:C6	C6

C1 is not leading at attributes A1, A2, A3, A6 and A8; C3 is not leading at attributes A1 and A4 to A9; C4 is not leading at attributes A2 to A5, and A7 to A9; and C6 is not leading at attributes A1 to A7 and A9. Thus, in a consortium including these companies, these companies cannot lead activities related to the attributes in which they underperform. Rather, they can be considered for transfer of skills from leading companies in these attributes. The next step focuses on the optimist's approach.

Optimist's approach

Table 28. Data matrix

	A1	A2	A3	A4	A5	A6	A7	A8	A9
C1	57	11	23	63	64	51	60	15	62
C3	39	60	61	41	43	33	36	42	34
C4	65	38	35	53	19	67	53	15	57
C6	22	49	52	21	33	18	15	64	16
Max	65	60	61	63	64	67	60	64	62
Company	C4	C3	C3	C1	C1	C4	C1	C6	C1

C1 is the top performer on four attributes (A4, A5, A7, A8); C3 is top on two (A2, A3) and C4 also tops on two (A1, A6). As a result, C1 is leading. C3 and C4 are strong contenders, and C6 is in the competition by virtue of being a 'niche' on attribute A8.

Table 29. Median-guided matrix

	A1	A2	A3	A4	A5	A6	A7	A8	A9
C1	57	11	23	63	64	51	60	15	62
C3	39	60	61	41	43	33	36	42	34
C4	65	38	35	53	19	67	53	15	57
C6	22	49	52	21	33	18	15	64	16
Median	41	43.5	38	35	38	40.5	35.5	36.5	39
Below average performers	C3 C6	C1 C4	C1 C4	C6	C4 C6	C3 C6	C6	C1 C4	C3 C6
Above average performers	C1 C4	C3 C6	C3 C6	C1 C3 C4	C1 C3	C1 C4	C1 C3 C4	C3 C6	C1 C4

In attributes where a company performs below average, it cannot be used for that attribute while those on above average, a company could be considered for inclusion on the bases of that attribute.

Minimize regret

Table 30. Regret matrix

	A1	A2	A3	A4	A5	A6	A7	A8	A9
C1	8	49	38	0	0	16	0	49	0
C3	26	0	0	22	21	34	24	22	28
C4	0	22	26	10	45	0	7	49	5
C6	43	11	9	42	31	49	45	0	46

At an attribute where a company shows zero regret, the company should be considered on the bases of that attribute.

EOL

Table 31. Regret matrix with EOL

	A1	A2	A3	A4	A5	A6	A7	A8	A9	EOL
C1	8	49	38	0	0	16	0	49	0	17.78
C3	26	0	0	22	21	34	24	22	28	19.67
C4	0	22	26	10	45	0	7	49	5	18.22
C6	43	11	9	42	31	49	45	0	46	30.67

The minimum EOL is obtained at C1. Thereafter, the order on merit of the next list is C4, C3, C6.

Consortia formation

Pair-based consortia: The strongest companies in descending order are C1, C4, C3, C6.

3.4.5. Consortium C1:C4.

Table 32. Performances of C1 and C4

	A1	A2	A3	A4	A5	A6	A7	A8	A9
C1	57	11	23	63	64	51	60	15	62
C4	65	38	35	53	19	67	53	15	57

The correlation of C1 and C4 is 0.51. The proposed consortium has the attributes below:

Table 33. Consortium C1:C4 performance

	A1	A2	A3	A4	A5	A6	A7	A8	A9	Mean	Std dev	CV
C1:C4	65	38	35	63	64	67	60	15	62	52.11	18.29	0.35

The consortium shows an improved strength with mean 52.1 compared to strength 45.1 of C1 and 44.7 of C4. The CV = 0.35 of cluster C1:C4 is lower than those of its components (C1 has CV = 0.49; C4 has CV = 0.42). Hence, the cluster has improved strength and improved stability. Despite these developments, this consortium is deficient at attributes A2, A3 and A8 when compared to the possibilities of strengths from other companies discussed earlier. The assessment starts with consortium C3:C4.

Table 34. Performances of C3 and C4

	A1	A2	A3	A4	A5	A6	A7	A8	A9
C3	39	60	61	41	43	33	36	42	34
C4	65	38	35	53	19	67	53	15	57

The correlation of C3 and C4 is 0.1966. The proposed consortium follows:

Table 35. Consortium C3:C4 performance

	A1	A2	A3	A4	A5	A6	A7	A8	A9	Mean	Std dev	CV
C3:C4	65	60	61	53	43	67	53	42	57	60	8.85	0.15

The strength of proposed consortium is 60.0, which is higher than strength 43.2 of company C3 and 44.7 of C4. The CV = 0.15 of cluster C3:C4 is less than those of its components. Thus, this cluster also has improved strength and improved stability. Despite the strong points shown, this consortium is still suboptimal at attributes A4, A5 and A7 to A9 as compared to the possibilities of strengths from other companies. Next is consortium C1:C3.

Table 36. Performances of C1 and C3

	A1	A2	A3	A4	A5	A6	A7	A8	A9
C1	57	11	23	63	64	51	60	15	62
C3	39	60	61	41	43	33	36	42	34

The correlation of companies C1 and C3 is -0.7318. The proposed consortium has:

Table 37. Consortium C1:C3 performance

	A1	A2	A3	A4	A5	A6	A7	A8	A9	Mean	Std dev	CV
C1:C3	57	60	61	63	64	51	60	42	62	57.78	7.07	0.12

The strength of proposed consortium is 60.0, which is higher than strength 45.1 of company C1 and strength 43.2 of company C3. The CV = 0.12 of this cluster is lower than those of its components. This cluster too, has improved strength and improved stability. This consortium is still suboptimal at attributes A1, A6 and A8 as compared to the possibilities of strengths from other companies.

3.4.6. Consortium C3:C6.

Table 38. Performances of C3 and C6

	A1	A2	A3	A4	A5	A6	A7	A8	A9
C3	39	60	61	41	43	33	36	42	34
C6	22	49	52	21	33	18	15	64	16

The correlation of C3 and C6 is 0.7115. The proposed consortium has:

Table 39. Consortium C3:C6 performance

	A1	A2	A3	A4	A5	A6	A7	A8	A9	Mean	Std dev	CV
C3:C6	39	60	61	41	43	33	36	64	34	45.67	12.45	0.27

The strength of proposed consortium is 45.7. It is only slightly higher than strength 43.2 of C3 and 44.7 of C4. The CV = 0.27 of this new cluster is higher than CV = 0.24 of C3, but lower than CV = 0.42 of C4. Hence, the strength of this cluster cannot be said to be convincingly better while the stability has also not improved. In addition to the weaknesses shown, the cluster is also sub-optimal at attributes A1, A4 to A7 and A9. The next is consortium C4:C6.

Table 40. Performances of C4 and C6

	A1	A2	A3	A4	A5	A6	A7	A8	A9
C4	65	38	35	53	19	67	53	15	57
C6	22	49	52	21	33	18	15	64	16

The correlation of C4 and C6 is -0.8027. The proposed consortium has:

Table 41. Consortium C4:C6 performance

	A1	A2	A3	A4	A5	A6	A7	A8	A9	Mean	Std dev	CV
C4:C6	65	49	52	53	33	67	53	64	57	54.78	10.4	0.19

The new strength of proposed consortium is 54.8, which is a significant improvement of the individual components of the consortium since C4 has strength 44.7 and C6 has strength 32.2. The CV = 0.19 is lower than those of its components. Hence, the cluster has improved strength and improved stability. The cluster has strengths exposed, but still shows sub-optimality at attributes A2 to A5, A7 and A9. The next is consortium C1:C6.

Table 42. Performances of C1 and C6

	A1	A2	A3	A4	A5	A6	A7	A8	A9
C1	57	11	23	63	64	51	60	15	62
C6	22	49	52	21	33	18	15	64	16

The correlation of C1 and C6 is -0.8920. The proposed consortium has:

Table 43. Consortium C1:C6 performance

	A1	A2	A3	A4	A5	A6	A7	A8	A9	Mean	Std dev	CV
C1:C6	57	49	52	63	64	51	60	64	62	58	5.96	0.10

The strength of proposed consortium is 58.0, which is higher than strengths 45.1 of company C1 and 32.2 of C6. The CV = 0.10 of the cluster is lower than those of its components. Hence, the cluster has improved strength and improved stability. This consortium is sub-optimal at attributes A1 to A3 and A6. Thus the pairs of consortia cannot give an optimal solution. There can be consortia of more than two companies. The next discussion proceeds to cases of more than two companies. The above accounts lead to the possibilities of the consortia pairs: C1:C4; C3:C4; C1:C3; C3:C6; C4:C6; C1:C6.

Table 44. Summary of consortia

Consortium	Mean strength	CV of consortium	Correlation of members	Old mean strengths	CVs of individual companies
C1:C4	52.1	0.35	0.51	45.1:44.7	0.49:0.42
C3:C4	60.0	0.15	0.20	43.2:44.7	0.24:0.42
C1:C3	57.8	0.12	-0.73	45.1:43.2	0.49:0.24
C3:C6	45.7	0.27	0.71	43.2:32.2	0.24:0.57
C4:C6	54.8	0.19	-0. 80	44.7:32.2	0.42:0.57
C1:C6	58.0	0.10	-0. 89	45.1:32.2	0.49:0.57

Discussion of table.

Strengths of consortia formed ('mean strength' column) are higher than the strengths of the original individual components ('old mean strengths' column). Thus the new consortia are improvements of the original individual components constituting these consortia. On the coefficient of variations (CVs), CVs of the consortia ('CV of consortium' column) are almost all lower than the CVs of the original individual companies ('CVs of individual companies' column). The exception existed with cluster C3:C6 which had a lower CV from C3. A lower CV is more desirable than a higher one as a sign of superior stability. Therefore, the consortia are improvements of the original components.

Observations

One observation about the correlation of the consortium formed is that being low or negative does not anything imply regarding the strength of the consortium. Cluster C3:C4 had a low positive correlation while C1:C6 had negatively high correlated companies. The two companies formed a strong consortium. Almost all the consortia showed improved stabilities (lower CVs) and stronger than the original members. Cluster C3:C6 was less stable. Also, all the pairs showed to be sub-optimal as some attributes were still suboptimal. Based on the mean strength, the strongest consortium was C3:C4. However, this consortium was not the most stable according to CV. Cluster C1:C6 was the most stable, and second strongest according to mean strength. This consortium is not optimal because some of its attributes were outwitted by corresponding ones of other companies.

Table 45. Summary of consortia

	Mean strength	CV of consortium	Correlation of members
Mean strength	1		
CV of consortium	0.317126	1	
Correlation of members	0.34569	0.772541	1

The observation made is that possible consortia in paired cases are such that the correlation of consortium members showed a high positive correlation coefficient (> 0.5) with the CV between the same members. This trait can be investigated further in another study.

3.4.7. Top three consortia of pairs of companies. The strongest consortium C3:C4 has mean strength of 60.0. It was formed from companies that initially had a low positive correlation of only 0.20. This consortium is sub-optimal at the attributes A4, A5, A7, A8, A9. The second strongest consortium C1:C6 has mean strength of 58.0, formed from companies that initially had a high negative correlation of -0.89. This consortium is also sub-optimal. The attributes identified to be sub-optimal are A1, A2, A3, A6. The next strongest consortium C1:C3 has mean strength of 57.8, formed from companies that initially had a high negative correlation of -0.73. This consortium is also sub-optimal at attributes A1, A6, A8.

Consortia of more than two companies

The idea was to form a cluster of more than two companies. The consortium starting with cluster C1:C6 is inevitable since it was explained that C1:C3 can address the sub-optimality problem. Hence, the new consortium is C1:C3:C4.

Table 46. Performance of consortium C1:C6

C1:C6	58.0	0.10	-0. 89	45.1:32.2	0.49:0.57

Table 46. Performances of consortium C1:C6 and C4

	A1	A2	A3	A4	A5	A6	A7	A8	A9	Mean	Std dev	CV
C1:C6	57	49	52	63	64	51	60	64	62	58	5.96	0.10
C4	65	38	35	53	19	67	53	15	57	44.7	19.0	0.42

Table 47. Consortium C1:C4:C6 performance

	A1	A2	A3	A4	A5	A6	A7	A8	A9	Mean	Std dev	CV
C1:C4:C6	65	49	52	63	64	67	60	64	62	60.7	6.12	0.10

Strength of proposed consortium is 60.7. However, sub-optimality occurred at attributes A1 and A2. For these two attributes, C3 in particular, showed no regrets, and can be examined. The solution sought is a consortium that maximizes all the possible benefits and minimizes all the detriments to the level at which it is practically possible.

Construction of an optimal consortium

Table 48. Performances of consortium C1:C4:C6 and C3

	A1	A2	A3	A4	A5	A6	A7	A8	A9	Mean	Std dev	CV
C1:C4:C6	65	49	52	63	64	67	60	64	62	60.7	6.12	0.10
C3	39	60	61	41	43	33	36	42	34	43.2	10.39	0.24

Table 49. Consortium C1:C4:C6 and C3 performances

	A1	A2	A3	A4	A5	A6	A7	A8	A9	Mean	Std dev	CV
C1:C3:C4:C6	65	60	61	63	64	67	60	64	62	62.88	0.038	2.37

Consortium C1:C3:C4:C6 is optimal. It possesses all the maximum benefits in each attribute. Its performance shows an increased strength. It also has the smallest CV. Thus, the optimal consortium derived for this study is C1:C3:C4:C6.

5. Discussion

Each method was able to identify strong and weak companies as well as weak and strong consortia. However, no single method was able to provide an optimal consortium. Iterations and amalgamations of distance and hierarchical clustering algorithms were necessary to verify that the weak consortia identified were indeed weak, and that the strong ones were indeed strong. Only this dynamic approach could provide an optimal consortium.

Conclusion

The logical iterations and conglomeration of various methods showed consistency in identifying strong and weak consortia. The methodical approach resulted in a dynamic, efficient and effective result. This approach showed to be crucial in optimization of the ultimate consortium formed.

Recommendation

Care should be taken during formation of consortia or other partnerships aimed at delivering results. A consortium should not be formed from unsubstantiated or speculative standpoint. The study recommends that for the purpose of clustering, application of cluster analysis should combine several different methods, and allow logical iterations.

References

1. Bergman, B. (2009). Conceptualistic pragmatism: A framework for Bayesian analysis? *IIE Transactions*, 41, pp. 86-93.
2. Dhillon, I.S. & Modha, D.S. (2001). Concept decompositions for large sparse text data using clustering, *Machine Learning*, 42 (1), pp. 143-175.

3. Dolnicar, S. & Lazarevski, K. (2009). Methodological reasons for the theory/practice divide in market segmentation, *Journal of Marketing Management*, 25 (4), pp. 357-373.
4. Everitt, B.S., Landau, S. & Leese, M. (2001). *Cluster analysis*, 4th edition. London: Oxford University Press, Inc.
5. Fawcelt, T. (2006). An introduction to ROC analysis, *Pattern Recognition Letters*, 27 (8), pp. 861-874.
6. Joshi, G.S. (2010). Infrastructure development strategies for inclusive growth: India's Eleventh Plan, *Leadership. Manage. Eng*., 10 (2), pp. 65-72, April.
7. Kaufman, L. & Rousseeuw, P.J. (2005). *Finding groups in data: An introduction to cluster analysis*, 2nd edition. New York: Wiley.
8. Koti, M. (2006). Transformation of South African higher education institutions through mergers: The Eastern Cape case. Dissertation for MBA degree, Management College of South Africa, Durban, South Africa.
9. Kraskov, A., Stögbauer, H., Andrzejak, R.G. & Grassberger, P. (2003). Hierarchical clustering based on mutual information.
10. Larson, J.S., Bradlow, E.T. & Fader, P.S. (2005). An exploratory look at supermarket shopping paths, *International Journal of Research in Marketing*, 22 (4), pp. 395-414.
11. Limpert, E., Stahel, W.A. & Abbt, M. (2001). Log-normal distributions across the sciences: Keys and clues, *BioScience*, 51, pp. 341-352.
12. Perruchet, P. & Peereman, R. (2004). The exploitation of distributional information in syllable processing, *Journal of Neurolinguistics*, 17, pp. 97-119.
13. Powers, D.M.W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation, *Journal of Machine Learning Technologies*, 2 (1), pp. 37-63.
14. Seker, S.E., Altun, O., Ayan, U. & Mert, C. (2014). A novel string distance function based on most frequent characters, *International Journal of Machine Learning & Computers*, 4 (2), pp. 177-183.
15. Seo, J. & Shneiderman, B. (2002). Interactively exploring hierarchical clustering results, *Computer*, 35 (7), pp. 80-86.
16. Sharma, S.C. (2006). *Operation research: Theory of games and travelling root problem*. New Delhi: Discovery Publishing House.
17. Triantaphyllou, E. (2000). *Multi-criteria decision making: A comparative study*. Dordrecht, Netherlands: Kluwer Academic Publishers.
18. Weisstein, E.W. (2003). *CRC concise encyclopedia of mathematics*, 2nd edition. London: Chapman & Hall/CRC.