SMACOF Hierarchical Clustering to Manage Complex Design Problems with the Design Structure Matrix

Never Stand Still | UNSW Canberra | School of Engineering and Information Telecommunication

Li Qiao, Mahmoud Efatmaneshnik, Michael Ryan

{l.qiao,m.Efatmaneshnik,m.ryan}@adfa.edu.au

Capability Systems Centre

School of Engineering and IT

UNSW, Canberra

Capability Systems Centre
UNSW Canberra

What is the problem?

Motivation

- The development of systems in defence industry, e.g. aircraft or satellite, involves numerous engineering from different disciplines.
- Know more about the structure of the problem since the number of interactions among various sub-problems can lead to a significant level of complexity.
- Trace the structure of design problems, match that to the way designers tackle those problems.
- A better understanding of the way designers work and why take the actions we see.
- Quantitative analysis to obtain insights of the design problem (hidden pattern and structures).

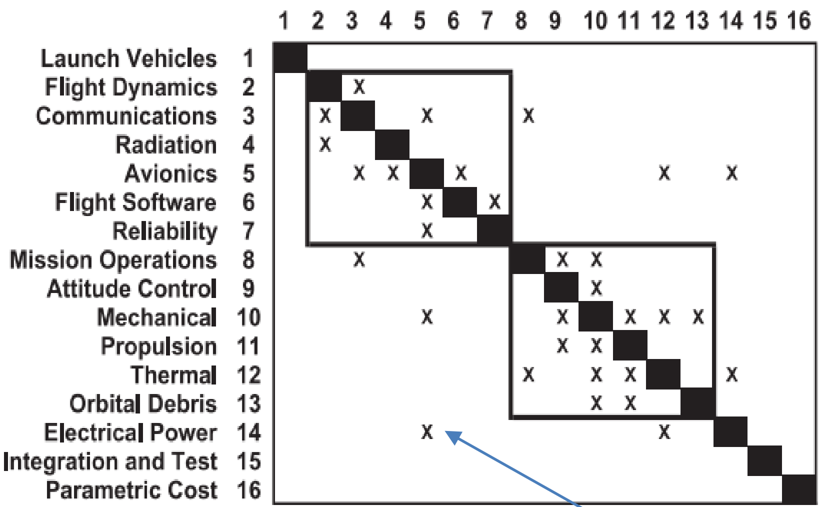
Background

- Design Structure Matrix is a matrix representation of a complex.
- Our previous work* regarding MDS clustering provides a rigorous study of its application to DSM modular analysis.
- Tailor our previous algorithm for the structuring of design problem.

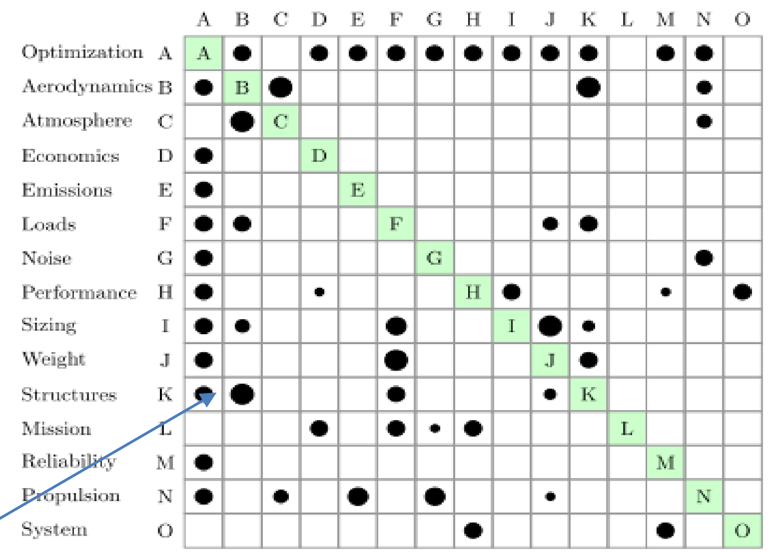
* Qiao, L., Efatmaneshnik, M., Ryan, M., Shoval, S.: Product modular analysis with design structure matrix using a hybrid approach based on MDS and clustering. J. Eng. Design **28**(6), 433–456 (2017)

Examples

Team-based design for satellite



Aircraft design problem



Notice the number of interactions among various sub-problems lead to a significant level of complexity.

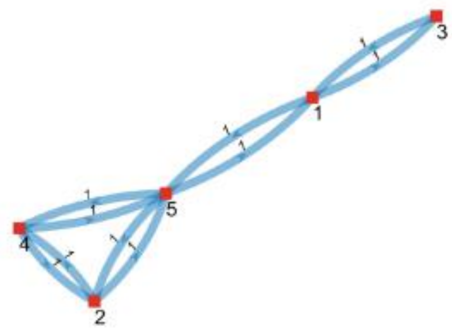
How to organize various subsystem design groups or their activities to ensure the efficiency of work?

How to do it?

Problem formulation

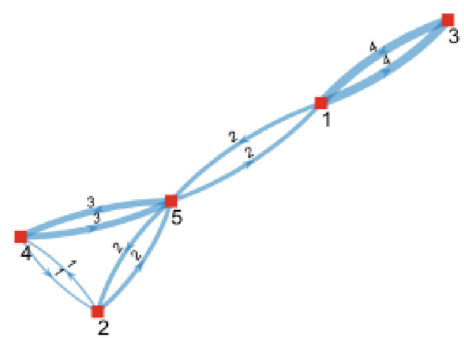
- DSM and graph representation of a design problem
  - A complex design problem can be represented by a graph of interactions.
  - The relationships between elements can be mapped to a matrix, called a DSM

	①	②	③	④	⑤
①			1		1
②				1	1
③	1				
④		1			1
⑤	1	1		1	



A binary DSM5
And its graph

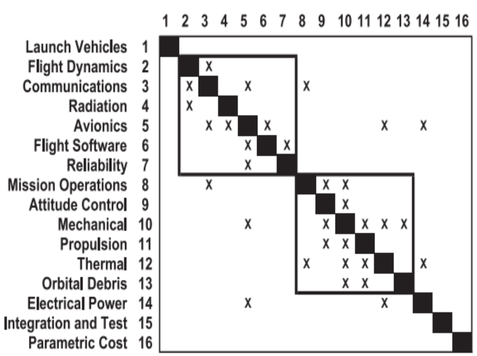
	①	②	③	④	⑤
①			4		2
②				1	2
③	4				
④		1			3
⑤	2	2		3	



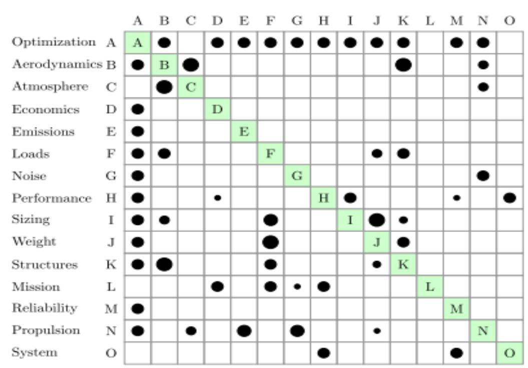
A integer DSM5
and its graph

Algorithm and Procedure

Algorithm	SMCOF hierarchical clustering with DSM
Input:	DSM (an $n \times n$ matrix)
Output:	Dendrogram, Partitions for a range of k with Cost, optimal partitions
Step 1: DSM and its digitalization	**If** the entries are not numbers, **then** digitize the DSM, **end if**.
Step 2: SMACOF	Use SMACOF, obtain embedded data (an $n \times m$ matrix)
Step 3: Hierarchical clustering	Compute the distance matrix using *Cosine*, hierarchical clustering the embedded data, obtain the partitions for a range of k
Step 4: Evaluation and knowledge deployment	Calculate the *Cost* for a range of k **If** the P_{ref} exits = True Calculate the *Jaccard* index between the P_{ref} and the obtained partitions **end** **If** there are no design constraints **then** choose the optimal partition when $k = k^*$ **elseif** there are design constraints, modify the partition **end if**



X→1



Assign an integer ranging from 1 to 5 to replace
The dots according to the size

Algorithm and Procedure

Algorithm	SMCOF hierarchical clustering with DSM
Input:	DSM (an $n \times n$ matrix)
Output:	Dendrogram, Partitions for a range of k with Cost, optimal partitions
Step 1: DSM and its digitalization	**If** the entries are not numbers, **then** digitize the DSM, **end if**.
Step 2: SMACOF	Use SMACOF, obtain embedded data (an $n \times m$ matrix)
Step 3: Hierarchical clustering	Compute the distance matrix using *Cosine*, hierarchical clustering the embedded data, obtain the partitions for a range of k
Step 4: Evaluation and knowledge deployment	Calculate the *Cost* for a range of k **If** the P_{ref} exits = True Calculate the *Jaccard* index between the P_{ref} and the obtained partitions **end** **If** there are no design constraints **then** choose the optimal partition when $k = k^*$ **elseif** there are design constraints, modify the partition **end if**

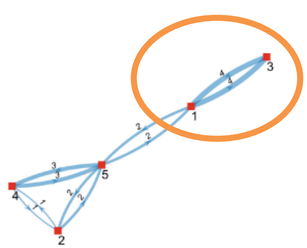
SMACOF: Scaling by Majorizing a Complicated Function is one type of multi-dimentinonal scaling (MDS)techniques.
Provide an approach for multidimensional scaling based on stress minimization by means for majorization
Represent elements in a few dimensional while preserving their multidimensional inter-interm similarity as close as possible.

Algorithm and Procedure

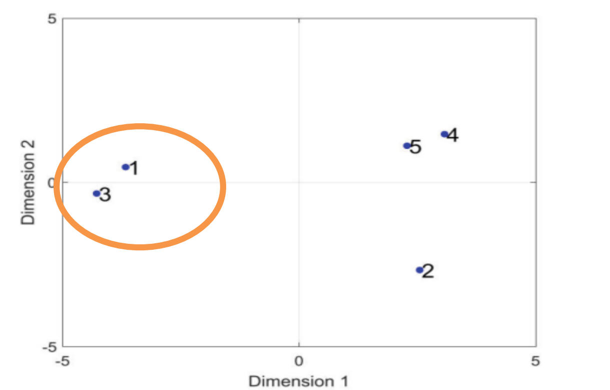
Algorithm	SMCOF hierarchical clustering with DSM
Input:	DSM (an $n \times n$ matrix)
Output:	Dendrogram, Partitions for a range of k with Cost, optimal partitions
Step 1: DSM and its digitalization	**If** the entries are not numbers, **then** digitize the DSM, **end if**.
Step 2: SMACOF	Use SMACOF, obtain embedded data (an $n \times m$ matrix)
Step 3: Hierarchical clustering	Compute the distance matrix using *Cosine*, hierarchical clustering the embedded data, obtain the partitions for a range of k
Step 4: Evaluation and knowledge deployment	Calculate the *Cost* for a range of k **If** the P_{ref} exits = True Calculate the *Jaccard* index between the P_{ref} and the obtained partitions **end** **If** there are no design constraints **then** choose the optimal partition when $k = k^*$ **elseif** there are design constraints, modify the partition **end if**

Input is DSM5 a 5×5 matrix

Set $m = 2$, output is a 5×2 matrix.



	①	②	③	④	⑤
①			4		2
②				1	2
③	4				
④		1			3
⑤	2	2		3	



Representation in new 2D by SMACOF

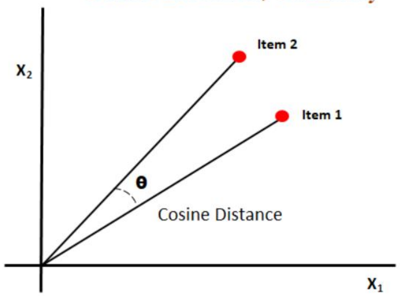
Algorithm and Procedure

Algorithm	SMCOF hierarchical clustering with DSM
Input:	DSM (an $n \times n$ matrix)
Output:	Dendrogram, Partitions for a range of k with Cost, optimal partitions
Step 1: DSM and its digitalization	**If** the entries are not numbers, **then** digitize the DSM, **end if**.
Step 2: SMACOF	Use SMACOF, obtain embedded data (an $n \times m$ matrix)
Step 3: Hierarchical clustering	Compute the distance matrix using *Cosine*, hierarchical clustering the embedded data, obtain the partitions for a range of k
Step 4: Evaluation and knowledge deployment	Calculate the *Cost* for a range of k **If** the P_{ref} exits = True Calculate the *Jaccard* index between the P_{ref} and the obtained partitions **end** **If** there are no design constraints **then** choose the optimal partition when $k = k^*$ **elseif** there are design constraints, modify the partition **end if**

Clustering is a type of unsupervised machine learning to discover the hidden structure.

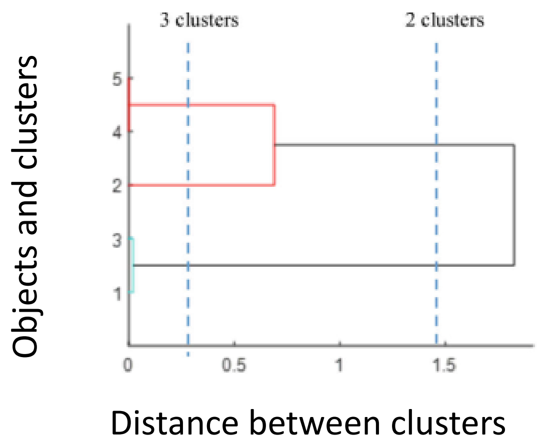
What we use is agglomerative hierarchical clustering: build a hierarchy structure of clusters by mearing all objects, producing a single, all-inclusive cluster at the top, and singleton clusters of individual objects at the bottom. Objects are compared using *Cosine* distance, the angular difference of the two data vectors. Two clusters with the smaller distance are joined.

Cosine Distance/Similarity



Algorithm and Procedure

Algorithm	SMCOF hierarchical clustering with DSM
Input:	DSM (an $n \times n$ matrix)
Output:	Dendrogram, Partitions for a range of k with Cost, optimal partitions
Step 1: DSM and its digitalization	**If** the entries are not numbers, **then** digitize the DSM, **end if**.
Step 2: SMACOF	Use SMACOF, obtain embedded data (an $n \times m$ matrix)
Step 3: Hierarchical clustering	Compute the distance matrix using *Cosine*, hierarchical clustering the embedded data, obtain the partitions for a range of k
Step 4: Evaluation and knowledge deployment	Calculate the *Cost* for a range of k **If** the P_{ref} exits = True Calculate the *Jaccard* index between the P_{ref} and the obtained partitions **end** **If** there are no design constraints **then** choose the optimal partition when $k = k^*$ **elseif** there are design constraints, modify the partition **end if**



The solutions of hierarchical clustering are called dendrogram.
Represent nested clusters for DSM5.
A quantitative description of data properties.
A partition can be obtained by cutting the dendrogram at a certain level.

$$P_{k=2} = \{\{1,3\},\{2,4,5\}\}$$

$$P_{k=3} = \{\{1,3\},\{4,5\},\{2\}\}$$

Algorithm and Procedure

Algorithm	SMCOF hierarchical clustering with DSM
Input:	DSM (an $n \times n$ matrix)
Output:	Dendrogram, Partitions for a range of k with Cost, optimal partitions
Step 1: DSM and its digitalization	**If** the entries are not numbers, **then** digitize the DSM, **end if**.
Step 2: SMACOF	Use SMACOF, obtain embedded data (an $n \times m$ matrix)
Step 3: Hierarchical clustering	Compute the distance matrix using *Cosine*, hierarchical clustering the embedded data, obtain the partitions for a range of k
Step 4: Evaluation and knowledge deployment	Calculate the *Cost* for a range of k **If** the P_{ref} exits $=$ True Calculate the *Jaccard* index between the P_{ref} and the obtained partitions **end** **If** there are no design constraints **then** choose the optimal partition when $k = k^*$ **elseif** there are design constraints, modify the partition **end if**

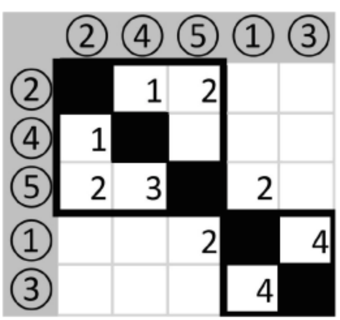
Compare the quality of the partition solutions using

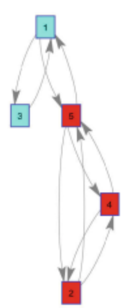
$$\begin{aligned} Cost &= \sum IntraClusterCost + \sum ExtraClusterCost \\ &= \underbrace{\sum [DSM(i,j) + DSM(j,i)] \times d_k}_{i,j \text{ are in the same cluster}} + \underbrace{\sum [DSM(i,j) + DSM(j,i)] \times n}_{i,j \text{ are not in the same cluster}} \end{aligned}$$

When we calculate the cost for a range of partitions, the cost value first decrease as the number of cluster increase. Then, the cost value increase, as a result of an increase in the number of interactions outside the partitions. The minimum value for the cost is considered as the optimal partition.

Algorithm and Procedure

Algorithm	SMCOF hierarchical clustering with DSM
Input:	DSM (an $n \times n$ matrix)
Output:	Dendrogram, Partitions for a range of k with Cost, optimal partitions
Step 1: DSM and its digitalization	**If** the entries are not numbers, **then** digitize the DSM, **end if**.
Step 2: SMACOF	Use SMACOF, obtain embedded data (an $n \times m$ matrix)
Step 3: Hierarchical clustering	Compute the distance matrix using *Cosine*, hierarchical clustering the embedded data, obtain the partitions for a range of k
Step 4: Evaluation and knowledge deployment	Calculate the *Cost* for a range of k **If** the P_{ref} exits = True Calculate the *Jaccard* index between the P_{ref} and the obtained partitions **end** **If** there are no design constraints **then** choose the optimal partition when $k = k^*$ **elseif** there are design constraints, modify the partition **end if**





$$P_{k=2} = \{\{1,3\},\{2,4,5\}\}$$ Cost = 72

$$P_{k=3} = \{\{1,3\},\{4,5\},\{2\}\}$$ Cost = 78

The optimal k^*=2. If there are no design constrains, We suggest the solution preference is $P_{k^*=2}$

Algorithm and Procedure

Algorithm	SMCOF hierarchical clustering with DSM
Input:	DSM (an $n \times n$ matrix)
Output:	Dendrogram, Partitions for a range of k with Cost, optimal partitions
Step 1: DSM and its digitalization	**If** the entries are not numbers, **then** digitize the DSM, **end if**.
Step 2: SMACOF	Use SMACOF, obtain embedded data (an $n \times m$ matrix)
Step 3: Hierarchical clustering	Compute the distance matrix using *Cosine*, hierarchical clustering the embedded data, obtain the partitions for a range of k
Step 4: Evaluation and knowledge deployment	Calculate the *Cost* for a range of k **If** the P_{ref} exits = True Calculate the *Jaccard* index between the P_{ref} and the obtained partitions **end** **If** there are no design constraints **then** choose the optimal partition when $k = k^*$ **elseif** there are design constraints, modify the partition **end if**

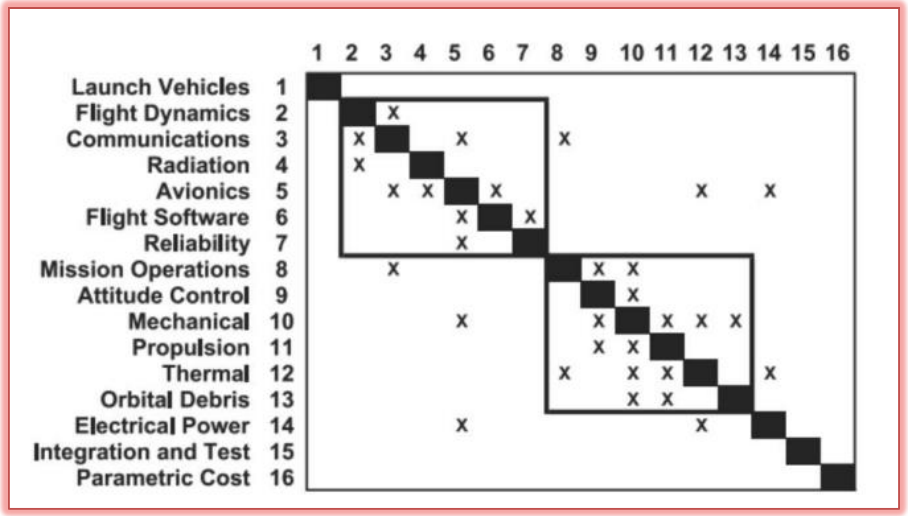
Use Jaccard index to compare partitions $Jaccard\ (X, Y) = |X \cap Y| / |X \cup Y|$

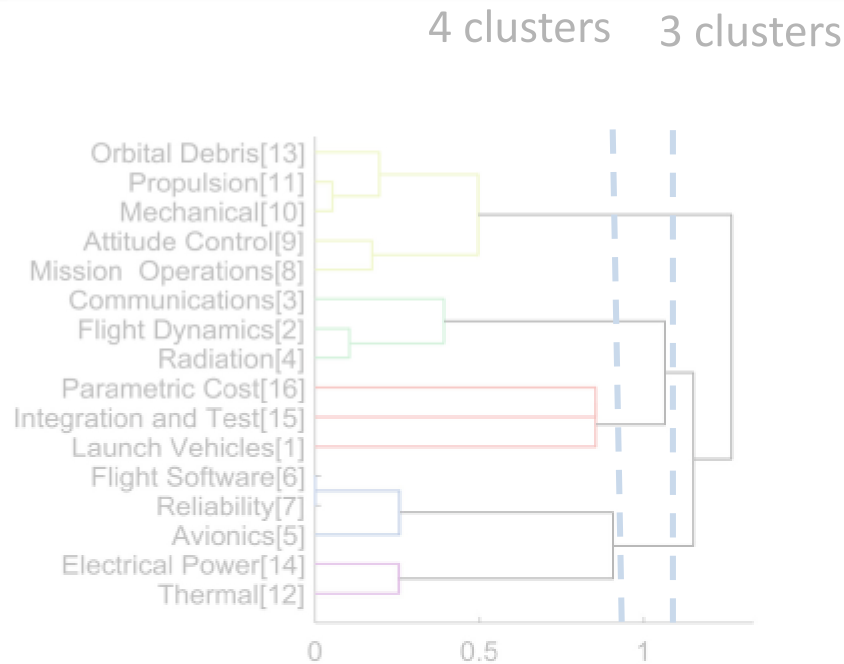
The number of members shared between both sets/the total number of members in both sets

1) Comparison of the obtained partitions with the reference partition P_{ref}
2) Comparison within the obtained partitions $P_2, P_3, \dots, P_{n-1}$

What are the results?

Industrial Case Study 1*





Team-based satellite design problem.
16 subsystems/disciplines, marked as DSM16
X presents the interaction among members
Transfer it to a binary matrix.
New-Girvan algorithm is adopted to obtain

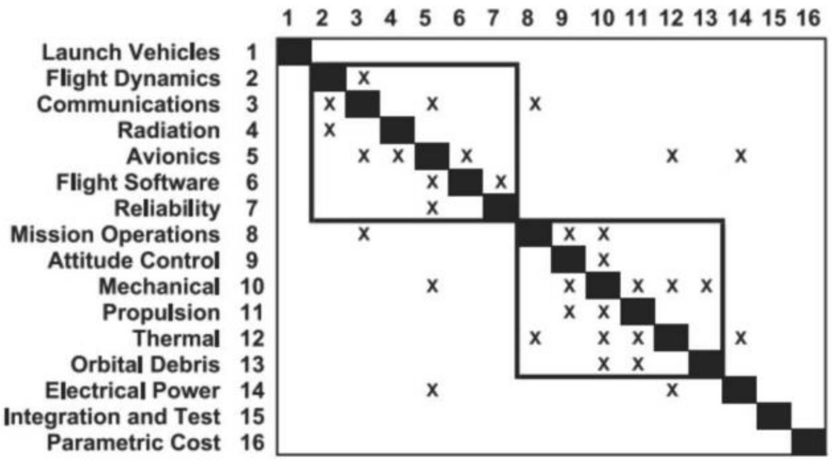
P_{ref}

$$\{\{1\},\{2,3,4,5,6,7\},\{8,9,10,11,12,13\},\{14\},\{15\},\{16\}\}$$

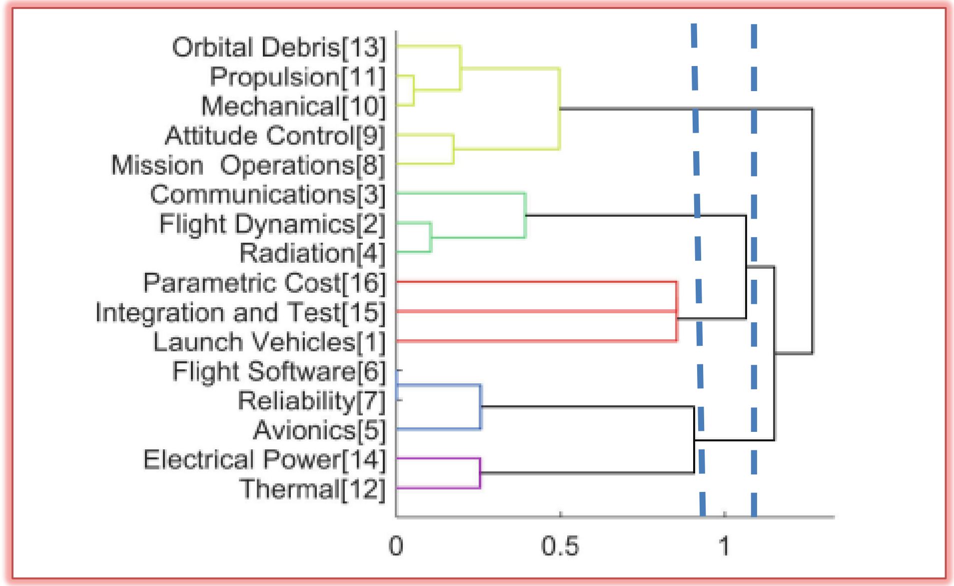
Dendrogram identify the related elements, such as Thermal[12] and Electrical power[14] .

* Avnet, M.S., Weigel, A.L.: An application of the design structure matrix to integrated concurrent engineering. Acta Astronautica **66**(5-6), 937–949 (2010)

Industrial Case Study 1*



4 clusters 3 clusters



Team-based satellite design problem.
16 subsystems/disciplines, marked as DSM16
X presents the interaction among members
Transfer it to a binary matrix.
New-Girvan algorithm is adopted to obtain

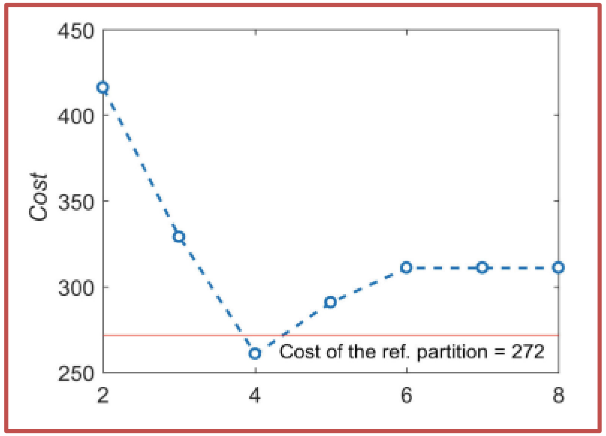
P_{ref}

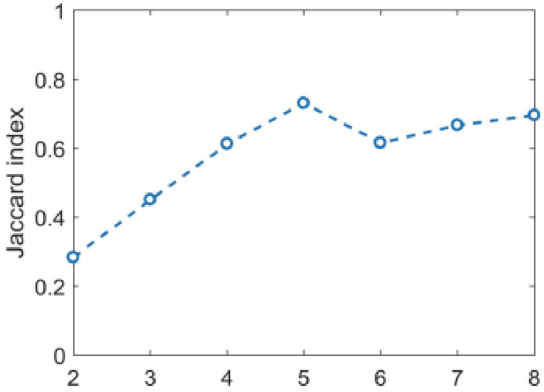
$$\{\{1\},\{2,3,4,5,6,7\},\{8,9,10,11,12,13\},\{14\},\{15\},\{16\}\}$$

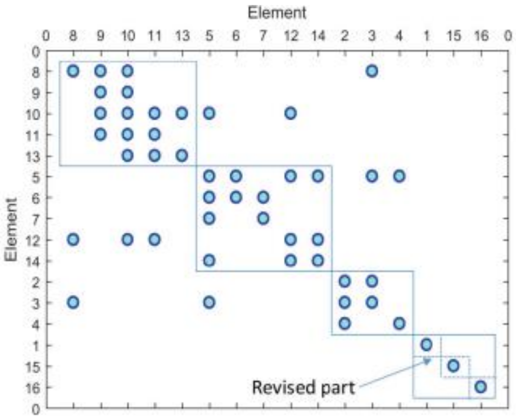
Dendrogram identify the related elements, such as Thermal[12] and Electrical power[14] .

* Avnet, M.S., Weigel, A.L.: An application of the design structure matrix to integrated concurrent engineering. Acta Astronautica **66**(5-6), 937–949 (2010)

Industrial Case Study 1*

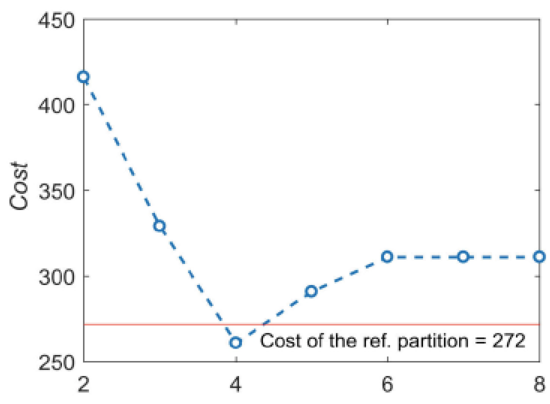


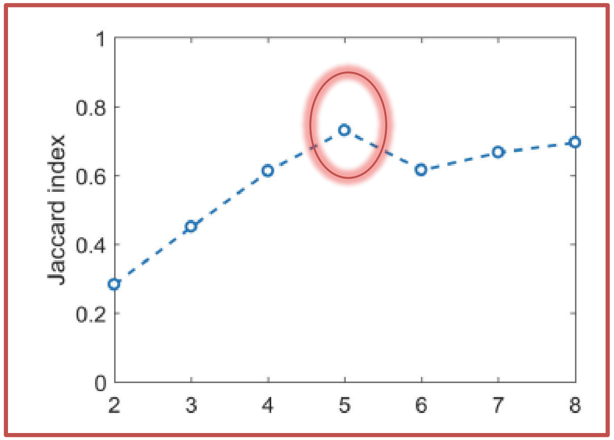


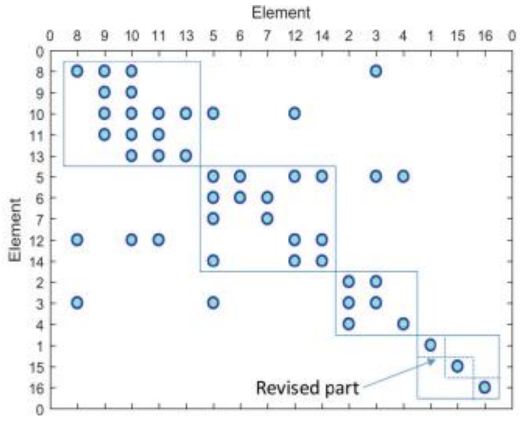


- Compare to P_{ref}, our result has lower cost value of 264.
- The optimal number of cluster $k^* = 4$.
- Compare the obtained 7 partitions ($P_{k=2}, \ldots, P_{k=8}$) to the P_{ref} with Jaccard index, $P_{k=5}$ is the most similar one to the reference partition.
- Launch Vehicle (1), Integration and Test (15) and Parametric Cost (16) are included though there are not any dependencies to other disciplines. Set the design constraints here that (1), (15) and (16) must be located in independent groups. Then we revise the partition with the same lowest cost

Industrial Case Study 1*

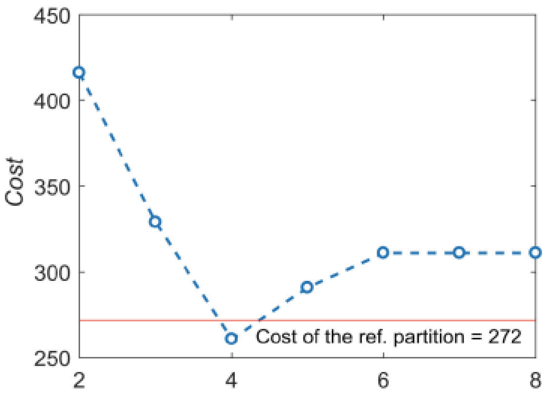


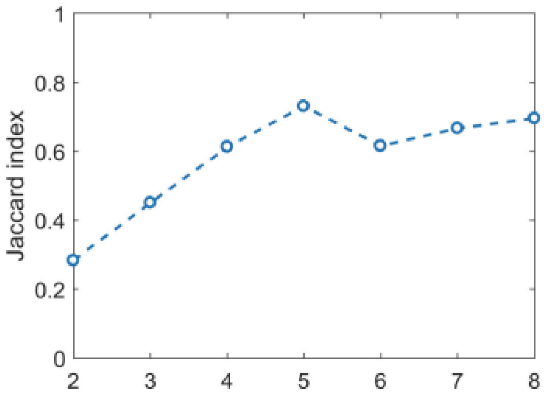


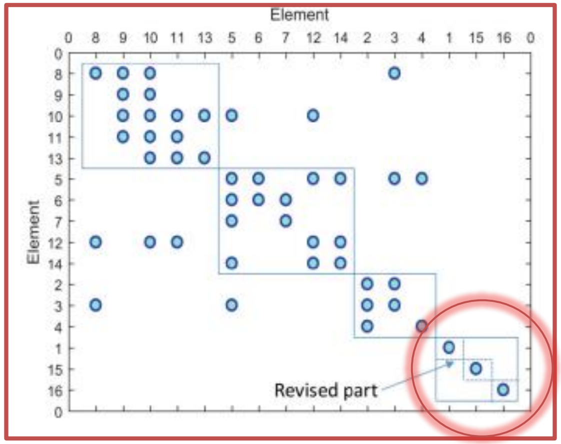


- Compare to P_{ref}, our result has lower cost value of 264.
- The optimal number of cluster $k^* = 4$.
- Compare the obtained 7 partitions ($P_{k=2}$,…, $P_{k=8}$) to the P_{ref} with Jaccard index, $P_{k=5}$ is the most similar one to the reference partition.
- Launch Vehicle (1), Integration and Test (15) and Parametric Cost (16) are included though there are not any dependencies to other disciplines. Set the design constraints here that (1), (15) and (16) must be located in independent groups. Then we revise the partition with the same lowest cost

Industrial Case Study 1*

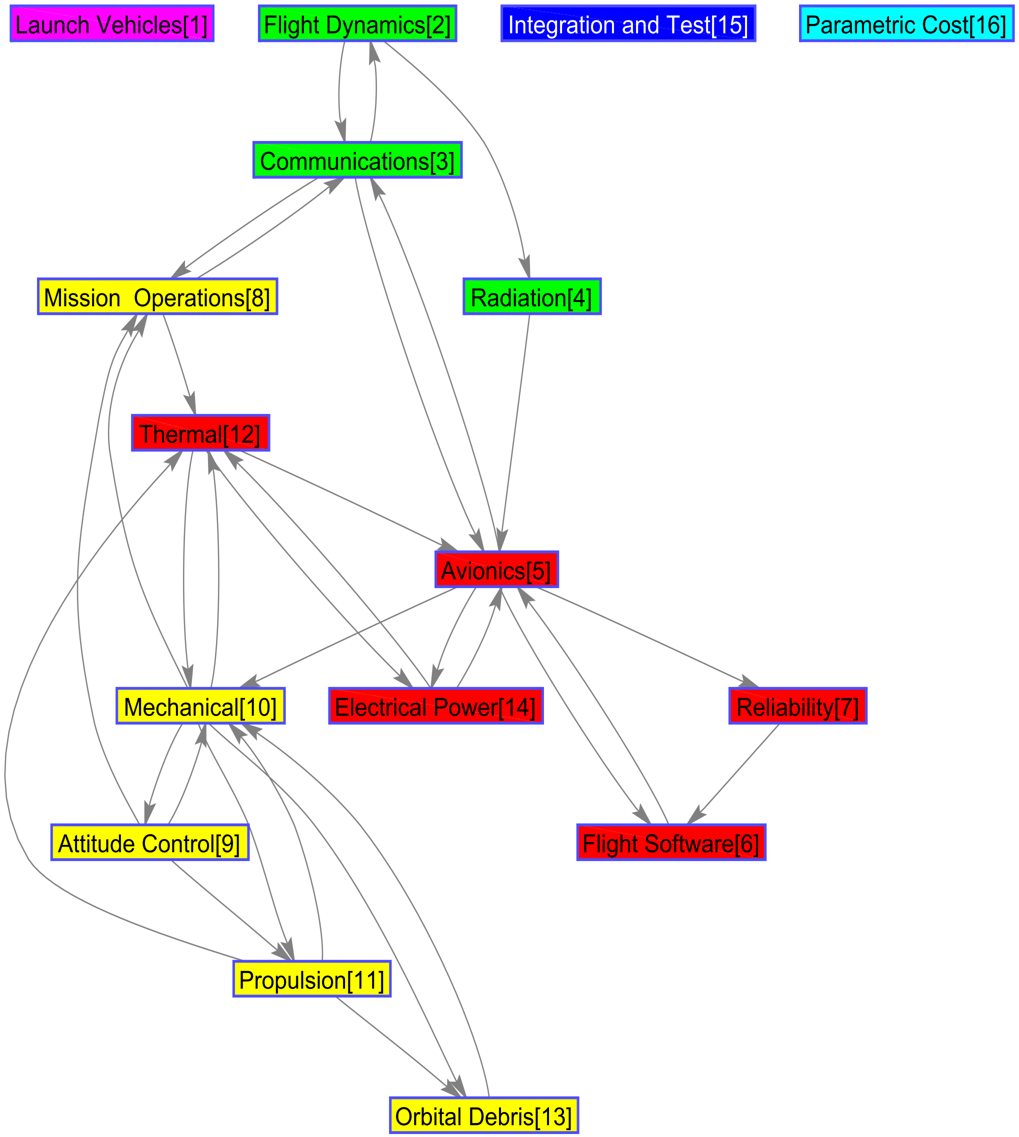






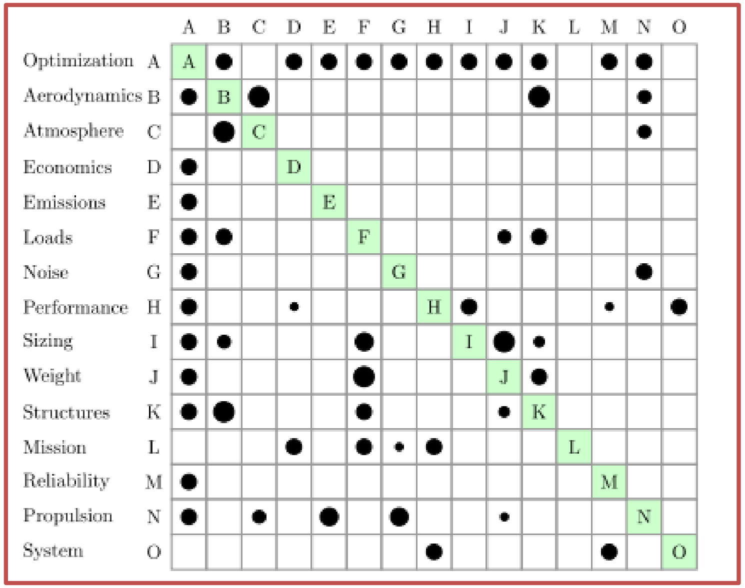
- Compare to P_{ref}, our result has lower cost value of 264.
- The optimal number of cluster $k^* = 4$.
- Compare the obtained 7 partitions ($P_{k=2}$,…, $P_{k=8}$) to the P_{ref} with Jaccard index, $P_{k=5}$ is the most similar one to the reference partition.
- Launch Vehicle (1), Integration and Test (15) and Parametric Cost (16) are included though there are not any dependencies to other disciplines. Set the design constraints here that (1), (15) and (16) must be located in independent groups. Then we revise the partition with the same lowest cost

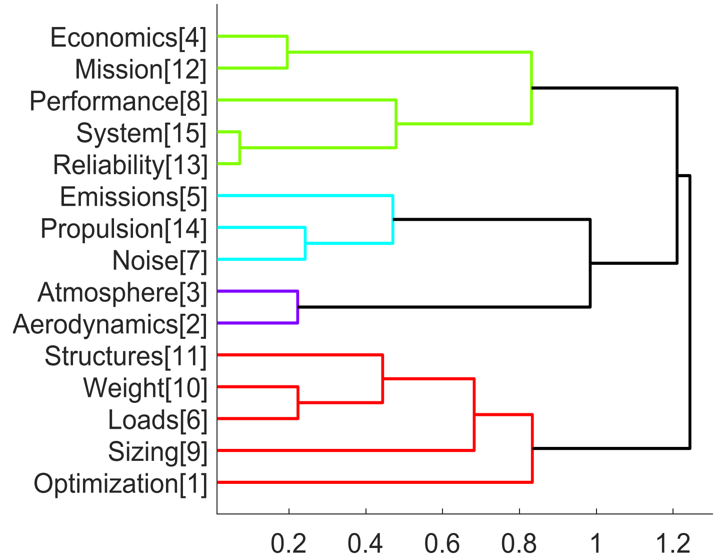
Industrial Case Study 1*



- Use a colour-coded graph to indicate the hierarchical modularity of the optimal partition.
- The teams with the same colour should work more closely together.

Industrial Case Study 2 *





Activity-based DSM15 for aircraft design problem.

15 elements, presenting various decision-making activities of the whole design problem.

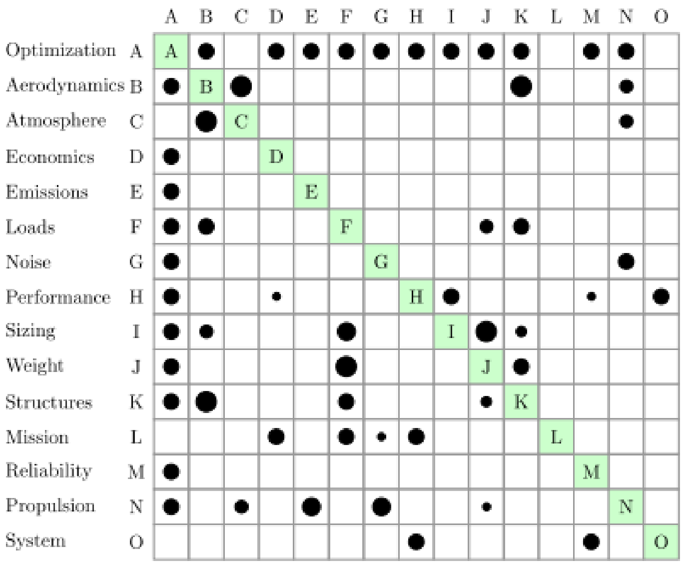
Entities represent the information flow between elements.

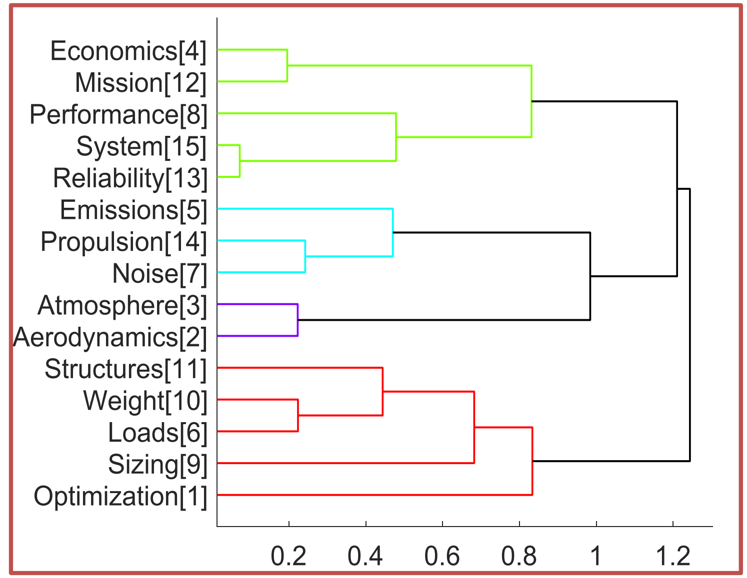
Larger dots denote stronger coupling between the disciplines.

The dendrogram represents nested clusters for DSM15.

* Lambe, A.B., Martins, J.R.R.A.: Extensions to the design structure matrix for the description of multidisciplinary design, analysis, and optimization processes. Struct. Multidisciplinary Optim. **46**(2), 273–284 (2012)

Industrial Case Study 2 *

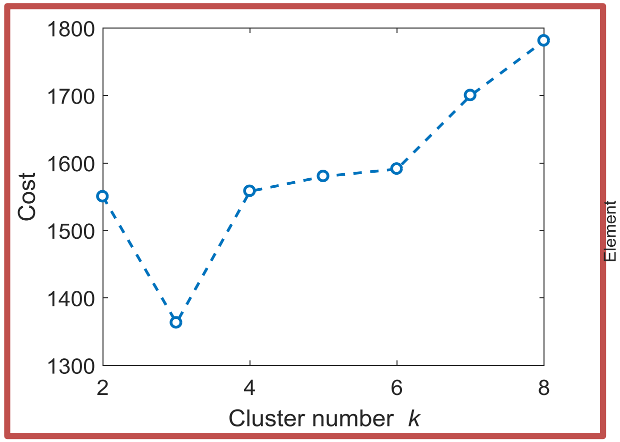


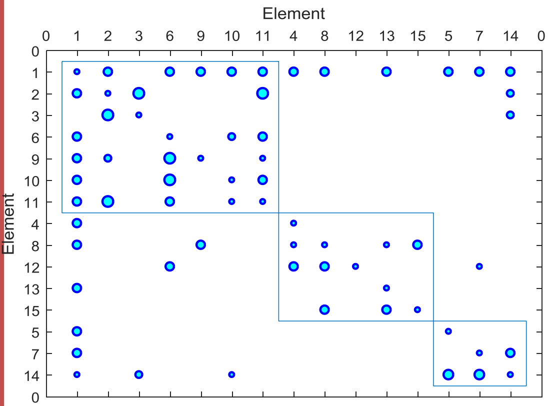


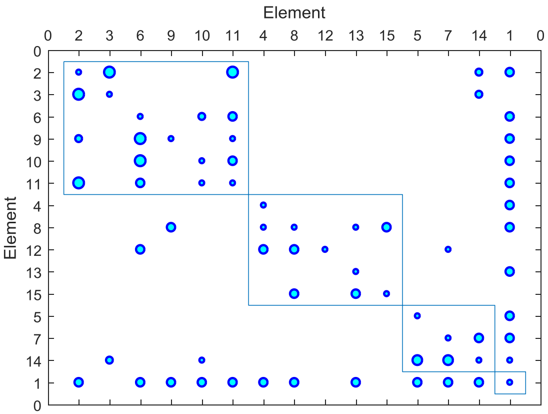
Activity-based DSM15 for aircraft design problem.
15 elements, presenting various decision-making activities of the whole design problem.
Entities represent the information flow between elements.
Larger dots denote stronger coupling between the disciplines.
The dendrogram represents nested clusters for DSM15.

* Lambe, A.B., Martins, J.R.R.A.: Extensions to the design structure matrix for the description of multidisciplinary design, analysis, and optimization processes. Struct. Multidisciplinary Optim. **46**(2), 273–284 (2012)

Industrial Case Study 2 *

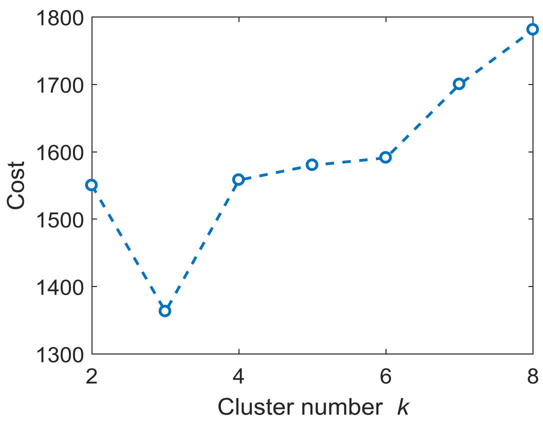


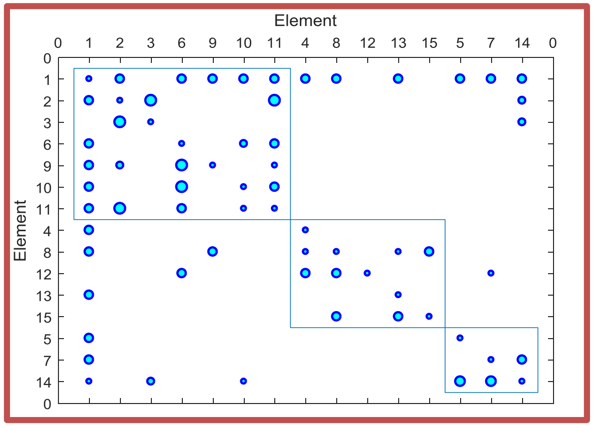


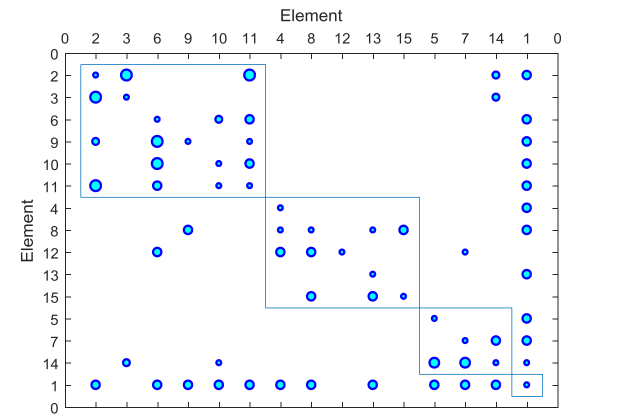


- Plot of cost vs.k suggests the optimal k^*=3.
- $P_{k=3}$ is not satisfied as element Optimization(1) is a bus-like element which has links to most of the rest activities.
- Look for a partition with low cost and in which the Optimization[1] is isolated.
- Revised our preferred partition to $P_{k=4}$.
- The rest elements group very well as very few interactions are outside the cluster.

Industrial Case Study 2 *

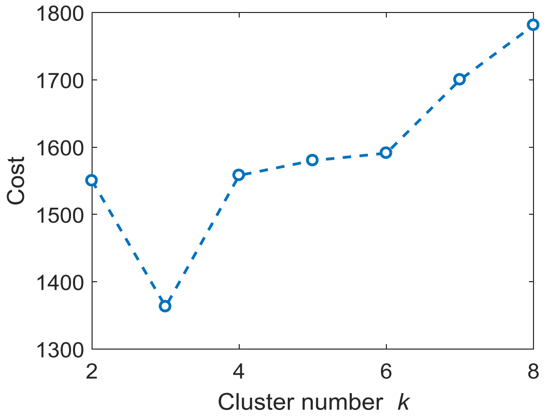


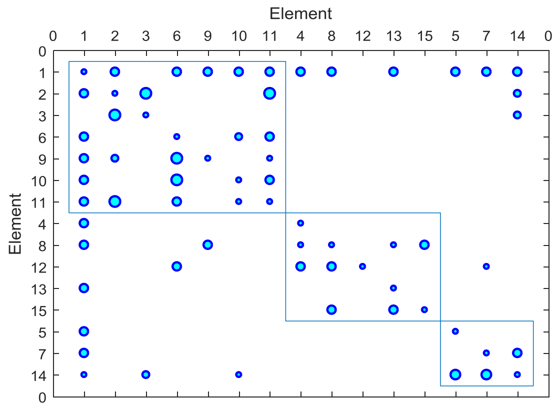


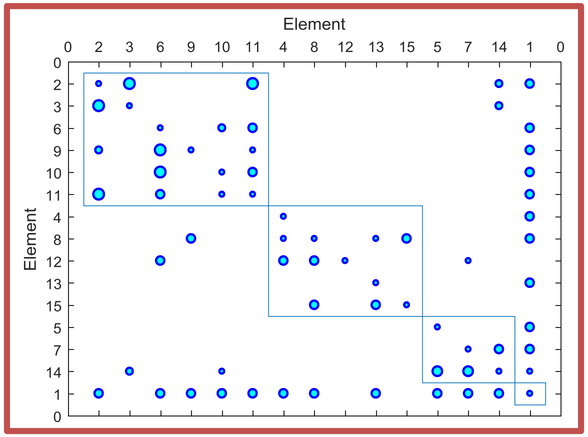


- Plot of cost vs.k suggests the optimal k^*=3.
- $P_{k=3}$ is not satisfied as element Optimization(1) is a bus-like element which has links to most of the rest activities.
- Look for a partition with low cost and in which the Optimization[1] is isolated.
- Revised our preferred partition to $P_{k=4}$.
- The rest elements group very well as very few interactions are outside the cluster

Industrial Case Study 2 *

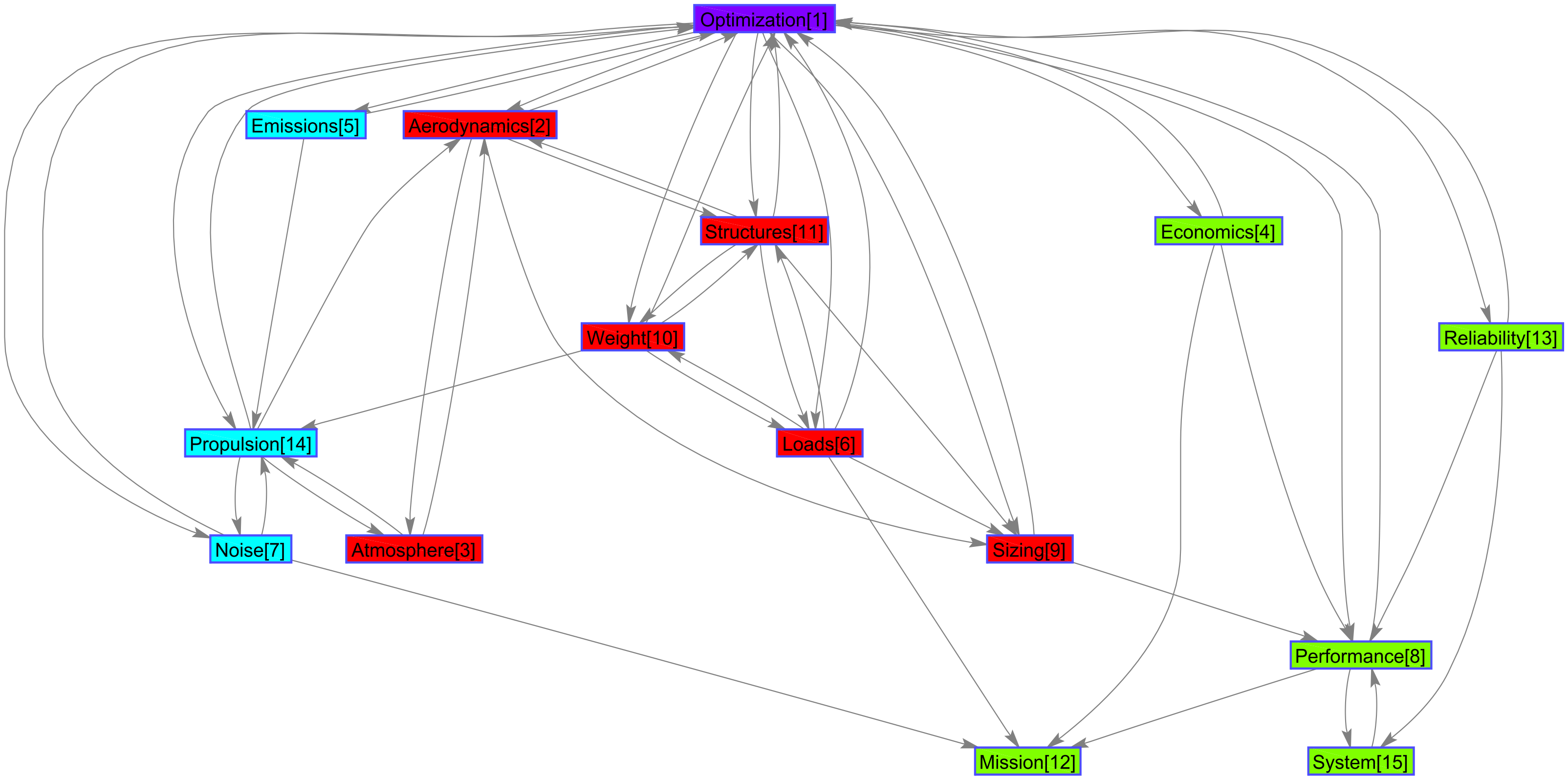






- Plot of cost vs.k suggests the optimal k^*=3.
- $P_{k=3}$ is not satisfied as element Optimization(1) is a bus-like element which has links to most of the rest activities.
- Look for a partition with low cost and in which the Optimization[1] is isolated.
- Revised our preferred partition to $P_{k=4}$.
- The rest elements group very well as very few interactions are outside the cluster

Industrial Case Study 2 *



- Use a colour-coded graph to indicate the hierarchical modularity of the optimal partition.
- The activities with the same colour should work more closely together.

Conclusion

Conclusion

Capability Systems Centre
UNSW Canberra

UNSW AUSTRALIA
Canberra

- Knowledge about the structure of a problem is important for deepening understanding of design.
- A DSM-based SMACOF hierarchical clustering method is proposed to model, analyse, and manage the complexity of design problems.
  - Perform well and capable to build a cluster hierarchy for a large problem.
  - Use *Cost* to determine the preference solution.
  - Use *Jaccard* to compare partitions and reveal more insights

Li Qiao, Mahmoud Efatmaneshnik, Michael Ryan
{l.qiao,m.Efatmaneshnik,m.ryan}@adfa.edu.au
UNSW, Canberra