



# Managing a complex project using a Risk-Risk Multiple Domain Matrix

Franck Marle, Catherine Pointurier, Hadi Jaber

► **To cite this version:**

Franck Marle, Catherine Pointurier, Hadi Jaber. Managing a complex project using a Risk-Risk Multiple Domain Matrix. *Journal of Modern Project Management*, 2015. <hal-01207083>

**HAL Id: hal-01207083**

**<https://hal.archives-ouvertes.fr/hal-01207083>**

Submitted on 30 Sep 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Managing a complex project using a Risk-Risk Multiple Domain Matrix

Catherine POINTURIER<sup>1</sup>, Franck MARLE<sup>2</sup>, Hadi JABER<sup>2</sup>

<sup>1</sup>CEA Centre DAM Ile-de-France, Arpajon, France

<sup>2</sup>Ecole Centrale Paris, Laboratoire Genie Industriel, Chatenay-Malabry, France

**Abstract:** This communication aims at presenting a clustering methodology applied to a complex project consisting of the delivery of three interdependent sub-systems. This enables small and complementary task forces to be constituted, enhancing the communication and coordination on transverse issues related to the complexity of the whole system. The problem is to gather and exploit data for such systems, with numerous and heterogeneous risks of different domains (product, process, organization). The method consists in regrouping actors through the clustering of the risks they own. The result is a highlight on important and transverse risk interdependencies, within and between projects. These should not be neglected in order to avoid potential severe issues, whether during the project or during the exploitation of its deliverable. An application on a real program of plant implementation in the CEA-DAM is presented, with a sensitivity analysis of the clustering results to the inputs and chosen configurations of the problem.

*Keywords:* Clustering, Risk interdependency, Complex Project Management, Multiple Domain Matrix

## 1 Introduction

This communication aims at presenting a clustering methodology applied to a complex project consisting of designing and implementing a new plant for a specific product related to CEA-DAM activity. CEA (*Commissariat à l'Energie Atomique*) is the French nuclear energy institution, and the DAM division (*Division des Applications Militaires*) is related to military applications. The project involves institutions on 5 different sites, with several divisions of the CEA and external contractors, with different cultures and tools. The constrained and uncertain nature of this kind of innovative project makes it obviously risky, with a strong importance of project risk management. The deliverable is structured into 3 Sub-Systems (SS1 to SS3), each of them contributing to the others with structuring intermediary deliverables, particularly the safety studies (due to the particular nature of raw material and outputs of this plant). The project is managed at a System level (SYST) and is exposed to external influences and risks (EXT). 5 categories are considered (SYST, EXT, SS1, SS2 and SS3), each of them including more or less project risks. The problem is to gather and exploit data for such systems, with numerous and heterogeneous risks of different domains (product, process, organization). The method consists in regrouping actors through the clustering of the risks they own. The result is a highlight on important and transverse risk interdependencies.

The remainder of the paper is as follows. Section 2 introduces the managerial and conceptual problem and the selected approach. Section 3 introduces how data are

gathered. Section 4 presents the clustering method, which is applied on the industrial project. Results are presented in Section 5 and discussed in Section 6.

## 2 Problem and research approach

The complexity of such projects is not only technical, but also and maybe essentially organizational. The number of actors involved, their diversity and the interdependence that potentially tie them with uncertainty on the occurrence and the strength of this relationship make it very difficult to manage such project organizations, both for managers or individual members.

The main risk is them to make locally optimal decisions, not considering the potential consequences for the rest of the project, generally because of ignorance of such potential consequences. Decisions are not well coordinated, propagation is not correctly anticipated, and finally the damages on the project will be higher, whether during the project (delays, overcosts...) or on the project deliverable (production performance, maintenance costs...). Several project elements can be considered for such studies, like product components, functions, activities or actors. Here, we use data about project risks to organize the relationships between the actors who own these risks.

Our aim is thus to develop a method to reshuffle project risk organization in order to put together (as much as possible) interconnected risks, and thus actors. There are three main blocks in our approach, corresponding to the following Sections of the paper: the data gathering, the customization of problem formulation and resolution strategy, and the resolution itself.

## 3 The data gathering

Two categories of relationships are particularly important in system modeling: hierarchical (vertical) and lateral (horizontal). Hierarchical relationships stem from the decomposition or breakdown of a system into smaller, more manageable elements. Hierarchical relationships are often modeled with breakdown structure diagrams.

We are here manipulating risks, which may be grouped in different levels of categories (or families), depending on their domain, their assessed values or ownership. The reliability of data related to risks and moreover risk interdependencies is already a question, but the conversion of such data into an assessment of pairs of actors interrelationships would add even more questions about the data reliability. For instance, compensation issues could arise from aggregation of several small interrelations, the addition of which becoming potentially higher than a single big one.

Lateral relationships stem from interactions between elements, such as flows of material or information, at the same level. While a DSM is mainly used to represent the lateral relationships between elements at a particular level of decomposition, it can also show elements locations in a hierarchy. The combinatory explosion of number of potential interdependencies between  $N$  risks is a major issue in such complex systems. There are  $N*(N-1)$  cells to fill or not, but each cell  $(i,j)$  has to be analyzed twice, once from  $R_i$  to  $R_j$  as an effect, and once from  $R_j$  to  $R_i$  as a cause. So, in reality, for  $N$  risks,  $2*N*(N-1)$  questions are to be answered.

The efficiency of the gathering process is crucial, particularly the sequence of the several pieces of analysis that have to be done. In this case, a direct assessment is proposed, simultaneous to the identification of a potential interdependency, which is new compared to previous applications. This is made possible by the use of an absolute scale, with several referring points permitting to identify properly the frontiers, the difference between one category and the other. This scale uses even numbers only, letting the possibility to use odd numbers in case of hesitation.

Table 1. Assessment scale

Interdependency numerical strength	Interdependency linguistic strength
10	Certain cause-effect link
8	The cause is very likely to trigger the effect
6	Likely to occur
4	Not likely to occur
2	Rare
0	Independent risks

We finally obtain 3 elements: a risk list, a Risk-Risk interdependency matrix, called **RR**, and an affiliation matrix, called **AR**.

**AR** is the responsibility assignment matrix of risks.  $AR_{ij}=1$  if Actor  $A_i$  is owner of Risk  $R_j$ , 0 else.

77 risks have been considered, from 5 categories (SubSystems 1 to 3, System, External). 495 potential interdependencies have been identified, the risk network being presented in Figure 1. It confirms the interest of using the DSM display since it is a compact way to represent systems interdependencies, whatever their level of complexity.

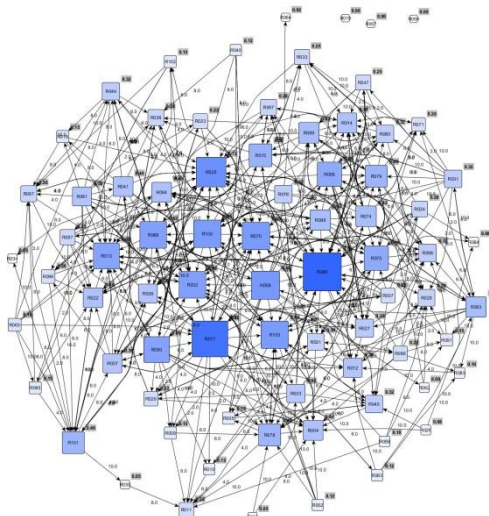


Figure1. Representationoftheexistingrisknetwork

The size of the nodes (and their color) is proportional to their centrality values (related to the number and weight of neighbor edges). Centrality analysis is not the object of this paper, but it is used as a visual indicator of the situation of risks within the current network and future clusters.

## **4 The resolution strategy**

This Section introduces the clustering algorithm used to group risks taking into account the number and strength of their interactions.

### **4.1 Related work on clustering**

Clustering is known as the identification of patterns around which communities of elements can be grouped (Gomez, Sanchez-Silva, & Duenas-Osorio, 2011). Specific DSM-related clustering techniques have been developed and implemented in industrial applications, like IGTA (Fredrik Borjesson & Katja Holta-Otto, 2012; Gutierrez-Fernandez, 1998; Idicula, 1995; Thebeau, 2001), or techniques designed for maximizing modularity in component DSMs (Helmer, Yassine, & Meier, 2010; Whitfield, Smith, & Duffy, 2002; Yu, Yassine, & Goldberg, 2007). On a general basis, clustering techniques are based whether on vertex similarity or on cluster fitness measure.

The first type of approaches is based on vertex similarity: the higher the vertex similarity, the stronger the need to cluster the vertices together. Clusters can be built progressively from singletons, or broken down from the initial graph into smaller clusters. These measures are based on a similarity matrix built from characteristics of the vertices. Rather than defining similarity measures, dissimilarity measures such as distance measures are usually defined, for instance the Euclidean and Manhattan distances (Ben-Arieh & Sreenivasan, 1999; Hennig & Hausdorf, 2006), or the Jaccard distance (Dong, Zhuang, Chen, & Tai, 2006) or the Levenshtein distance (Gusfield, 1997). The partitioning can be done without knowing the number of clusters  $k$  in advance, or requires this information like in the  $k$ -means method (McQuenn, 1967) and spectral clustering (Bühler & Hein, 2009). Some techniques have emerged to determine  $k$  or an interval for  $k$  (Tan, Broach, & Floudas, 2007). More generally, they are based on whether approximate heuristics or optimization algorithms, like (Sherali & Desai, 2005). A cluster can contain identical or similar elements, with a particular element called centroid and representative of the group (Filippone, Camastra, Masulli, & Rovetta, 2008). On the opposite, some works thus focus on edges that are least central or most “between” clusters, and remove them from the original graph in order to build the strongest clusters with the remaining edges (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008; Freeman, 1977; Girvan & Newman, 2002). Newman, Clauset and Girvan are co-authors of numerous works in the field of finding community structures in complex networks (Clauset, Moore, & Newman, 2006; Clauset, Newman, & Moore, 2004; Leicht & Newman, 2008; Newman & Web, 2003; Newman, 2006).

On the other hand, some clustering processes are based on cluster fitness measures, that is to say functions which assess the overall quality and relevance of a given cluster or of a given global clustering solution. The global objective of these methodologies is to identify clustering solutions which directly fulfill a certain property. For instance, methodologies based on graph density measures have been developed in order to

partition the initial graph into sub graphs, the density of which should be inferior and/or superior to chosen values (Karp, 1976; Kim, 2003; Zotteri, Kalchschmidt, & Caniato, 2005). But other cluster fitness measures are used as a criterion for graph partitioning. Indeed, as noticed by Schaeffer (2007), “one measure that helps to evaluate the sparsity of connections from the cluster to the rest of the graph is the cut size. The smaller the cut size, the better isolated the cluster”. Indeed, cut size-based measures undoubtedly permit to quantify the relative independence of a sub graph to the rest of the graph and have been used in many clustering processes (Kannan, Vempala, & Vetta, 2001; Shi & Malik, 2000). Finding the partition which minimizes cut-sizes (with restriction conditions on the orders of the sub-graphs) makes it possible to maximize the sum of the weights of the edges which are internal to the clusters. This cut-based measure is of great interest for our case. In order to facilitate complex project management, the reduction of interfaces in terms of number, and above all strength, is likely to be desirable.

## 4.2 Customizing the Resolution Strategy

The approach aims at maximizing interactions within risk clusters, and proposes risk owner groups corresponding to risk clusters. The main originality of this paper is to combine several results obtained from different clustering algorithms running with different problem configurations.

### 4.2.1 The optimization function

$\mathbf{RC}$  is a  $NR \times NC$  variable matrix with each of its elements  $RC_{j,k}$  ( $1 \leq j \leq NR$ ,  $1 \leq k \leq NC$ ) being a Boolean variable. For each risk, the variable  $RC_{j,k}$  being 1 means the presence of Risk  $R_j$  in Cluster  $C_k$ , while being zero means its absence.  $\mathbf{RC}$  is our decision variable. The objective function of the problem is defined by the sum of the values of all interactions between risks which belong to a same cluster. It is a quadratic integer problem, described in Eq. 1:

$$\max \sum_{1 \leq k \leq NC} \sum_{1 \leq j_1, j_2 \leq NR} RC_{j_1, k} * RC_{j_2, k} * RR_{j_1, j_2} \quad (1)$$

Figure 2 shows how reshuffled affiliation matrix  $\mathbf{AC}$  is deduced from risk clusters ( $\mathbf{RC}$ ) and risk ownership ( $\mathbf{AR}$ ) matrices.

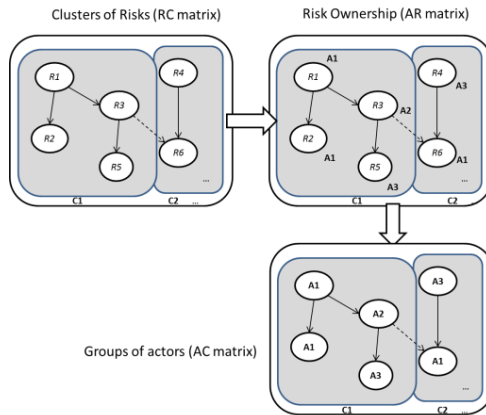


Figure 2. Actor groups are deduced from risk clusters and risk ownership

#### 4.2.2 The pre-treatment: defining the parameters

One key issue is to specify optimization problem parameters, both in terms of constraints and in terms of algorithm. Several constraints are added to the problem formulation, like maximal Cluster Size and maximal Number of Actors per Cluster. They allow clusters to be more manageable, but of course they reduce the capacity to bring together more interrelated elements. But it is very difficult for the decision-maker to specify *a priori* such parameters. This is why a range is defined for each parameter, and several scenarios are built considering different values within this range for all parameters. The interesting cases are when the risks are never clustered together and when they are always proposed in the same cluster. This can give an indication on the robustness of the decision to put them together, or to keep them in separated clusters.

#### 4.2.3 The treatment: combining the results of 4 algorithms

Instead of examining the quality of the clustering results produced by several algorithms, we decided to assemble a solution from different pieces of solutions given by different algorithms and problem configurations. This means that we did not choose an algorithm, but we choose to try to extract the best from each of them. Our resolution strategy is thus based on 4 well-known algorithms:

- Two algorithms based on modularity optimization (Blondel et al., 2008; Leicht & Newman, 2008),
- One DSM – clustering algorithm (Fredrik Borjesson & Katja Holtta-Otto, 2012)
- One spectral algorithm (Bühler & Hein, 2009).

This benchmark involved a testing process, and all the algorithms were applied to an existed risk network with specific community structure (Marle & Vidal, 2014); the result was suitable to the known structure.

#### 4.2.4 The post-treatment: considering some specific risks

Moreover, some risks may have numerous interdependencies, meaning that it is difficult to put them into a single cluster. They can be considered separately, whether by assigning them to more than one cluster, or by putting them out of the clusters, meaning that they are transverse and connected to every cluster. This is done here as a post-treatment action, considering clusters proposed by several algorithms using different problem configurations.

### 5 Results

For confidentiality reasons, no name is given, only Ids, both for sub systems, risks and actors. The results of the scenarios run have been analyzed and compared, in terms of performance (how many interactions are put within clusters, what is the size of clusters and consequently the size of human groups) and similarity (how many times are these risks proposed in the same clusters).

Several clusters have been picked up from different scenarios (supervised or unsupervised, with small groups or bigger and less dense ones, allowance of non-disjunction between clusters, consideration of several specific risks). This has enabled the global solution to be improved and to be closer from managerial reality.

The final version of proposed clusters is presented hereunder on Figure 3.

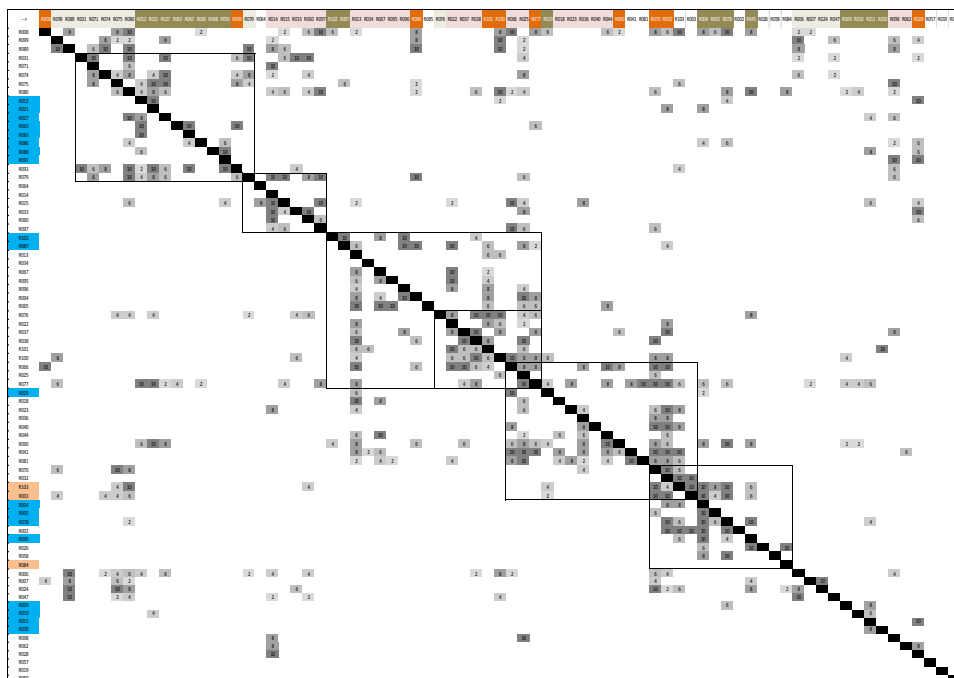


Figure3. Final clusters configuration based on combination of different solutions



Due to the difficulty to read such matrices, some extracts are showed hereunder. Clusters are numbered from the top-left (C1) to the bottom-right (C6). What is noticeable on a technical point of view is that:

- Although the density of the matrix is substantially higher than those obtained in previous industrial applications (8% instead of 3 to 4%), only 3 risks have been proposed as transverse risks (on the top-left corner of Figure 3), outside any cluster but related to all,
- Although some risks are big hubs (they have numerous inputs and outputs), they have always been proposed in a single cluster, with consistent neighbors. No inadequate separation has been proposed.
- C4 is a sub-cluster of C3, meaning that it is completely included into a bigger cluster proposed by another algorithm. This means that two kinds of meetings can be run, with an extended version of the task force in the case of C4.
- Once disjunction constraint has been relaxed, several chunks have been proposed, putting 1 to 3 risks into two clusters simultaneously. No other and more complicated situation was present (the same risk into 3 or more clusters, or risks which could not assure the readability of the matrix).

Once technical sanity check has been performed, the managerial interpretation of results has been done. This is the object of the following Section.

## **6Discussion**

Several points are noticeable on a managerial point of view.

Cluster C1 has proposed a combination of actors from Sub Systems 1 and 2, which was exactly what was required to improve the coordination and integration of these elements. It has been decided to implement it.

Cluster C2 is focused on SS 3 design including one risk related to SS 1 and one external risk. Cluster 3 is also mainly related to SS 3, but more focused on the exploitation phase. The Cluster 4 which is a sub-cluster of C3 but obtained with other algorithms serves as a kind of steering committee for interfacing SS3 design with the rest of the project. Namely, risks of C4 are related to SS1 (2), SS3 (5) and SYS (3). It is to be noticed also that the part of C3 which is not C4 contains 5 risks from SS3 but also 2 from SS2 and 1 from EXT, meaning that even a SS3 design-centered cluster contains diverse risks from other parts of the project.

Cluster C5 is globally centered on SS3 installation, but with a cost-oriented vision, rather than technical-oriented. This means that several risks come from SS3 and several others come from SYS (the system level) or EXT (the external factors that could affect costs). Finally, C6 ties SS2 to high-level events or decisions related to SYS and EXT.

Globally, the decision-maker found managerial relevance to all proposed clusters. This is due to the construction mode of the solution, since it has been built from different solutions obtained from different algorithms, meaning that the decision-maker validated the inclusion of each cluster in the solution.

The most important point is that it has been decided to implement the proposed groups, respectively containing 9, 3, 8, 6, 8 and 6 actors for clusters C1 to C6. Particularly, C4 is a 9-risk cluster with 6 different actors, and C3 is a 18-risk cluster containing C4, but with only two additional actors. But the proposal of running two different series of meetings has been kept. The agenda is broader in the case of C3.

Extensions to this work are to help to data gathering efficiency (the effort and the reliability of data if obtained from different sources) and to post-treatment strategy definition. The manual extraction of clusters or pieces of clusters from different scenarios is the main limit of the approach. The heart of the resolution, the algorithms, is not the key issue in the global approach, even if processing time can be limited. In this case, the highest calculation time for a scenario was 2 hours, meaning that it was still possible to run several scenarios for this given algorithm.

## References

- Ben-Arieh, D., & Sreenivasan, R. (1999). Information Analysis in a Distributed Dynamic Group Technology Method. *International Journal of Production Economics*, 60-61, 427–432.
- Blondel, V., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, P10008.
- Bühler, T., & Hein, M. (2009). Spectral clustering based on the graph p-Laplacian. In *Proceedings of the 26th Annual International Conference on Machine Learning* (pp. 81–88). ACM.
- Clauset, A., Moore, C., & Newman, M. (2006). Structural Inference of Hierarchies in Networks. In *Proceedings of the 23rd International Conference on Machine Learning*. Pittsburgh, PA.
- Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. *Phys. Rev. E*, 70(066111), 1–6.
- Dong, Y., Zhuang, Y., Chen, K., & Tai, X. (2006). A hierarchical clustering algorithm based on fuzzy graph connectedness. *Fuzzy Sets and Systems*, 157, 1760–1774.
- Filippone, M., Camastra, F., Masulli, F., & Rovetta, S. (2008). A survey of kernel and spectral methods for clustering. *Pattern Recognition*, 41.
- Fredrik Borjesson, & Katja Holttä-Otto. (2012). IMPROVED CLUSTERING ALGORITHM FOR DESIGN STRUCTURE MATRIX. In *ASME 2012 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference Physical Review*.
- Freeman, L. (1977). Set of measures of centrality based on betweenness. *Sociometry*, 40, 35–41.
- Girvan, M., & Newman, M. (2002). Community structure in social and biological networks. *Proceedings of National Academy of Science (PNAS)*, 99(12), 7821–7826. doi:doi:10.1073/pnas.122653799
- Gomez, C., Sanchez-Silva, M., & Duenas-Osorio, L. (2011). Clustering methods for risk assessment of infrastructure network systems. *Applications of Statistics and Probability in Civil Engineering*, 1389–1397.
- Gusfield, D. (1997). *Algorithms on Strings, Trees and Sequences*. Cambridge University Press.
- Gutierrez-Fernandez, C. I. (1998). *Integration Analysis of Product Architecture to Support Effective Team Co-Location*. Massachusetts Institute of Technology.
- Helmer, R., Yassine, A., & Meier, C. (2010). Systematic Module and Interface Definition using Component Design Structure Matrix. *Journal of Engineering Design*, 21(6), 647–675.
- Hennig, C., & Hausdorf, B. (2006). A Robust Distance Coefficient Between Distribution Areas Incorporating Geographic Distances. *Systematic Biology*, 55(1), 170–175.
- Idicula, J. (1995). *Planning for Concurrent Engineering*. Gintic Institute, Singapore.

- Kannan, R., Vempala, S., & Vetta, A. (2001). *On Clusterings : Good, Bad and Spectral. Working paper.*
- Karp, R. M. (1976). Probabilistic analysis of partitioning algorithms for the travelling salesman problem in the plane. *Mathematics of Operations Research*, 2(3), 209–224.
- Kim, S. (2003). Graph theoretic sequence clustering algorithms and their applications to genome comparison. In J. T. L. Wu, C.H., Wang, P., Wang (Ed.), *Chapter 4 in Computational Biology and Genome In-formatics* (in Wu, C.H.). Singapore: World Scientific.
- Leicht, E. A., & Newman, M. E. (2008). Community structure in directed networks. *Physical Review Letters*, 100(11), 118703.
- Marle, F., & Vidal, L. (2014). Forming risk clusters in projects to improve coordination between risk owners. *Journal of Management in Engineering*.
- McQuenn, J. (1967). Some methods for classification and analysis of multivariate observations. *Computers and Chemistry*, 4, 257–272.
- Newman, M. E. J. (2006). Finding community structure in networks using the eigenvectors of matrices. *Proc. Natl. Acad. Sci. USA*, 103, 8577–8582.
- Newman, M. E. J., & Web, W. (2003). Properties of highly clustered networks, 1–7.
- Sherali, H., & Desai, J. (2005). A global optimization RLT-based approach for solving the hard clustering problem. *Journal of Global Optimization*, 32(2), 281–306.
- Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8), 888–905.
- Sosa, M., & Marle, F. (2013). Assembling Creative Teams in NPD Using Creative Team Familiarity. *Journal of Mechanical Design*, 135.
- Tan, M. P. M., Broach, J. R. J., & Floudas, C. a. C. (2007). A novel clustering approach and prediction of optimal number of clusters: global optimum search with enhanced positioning. *Journal of Global Optimization*, 39(3), 323–346. doi:10.1007/s10898-007-9140-6
- Thebeau, R. E. (2001). *Knowledge Management of System Interfaces and Interactions for Product Development Process*. Massachusetts Institute of Technology.
- Whitfield, R., Smith, J., & Duffy, A. (2002). Identifying Component Modules. In *Proceedings of the 7th international conference on artificial intelligence in design AID'02* (pp. 571–592).
- Yu, T. L., Yassine, A., & Goldberg, D. (2007). An Information Theoretic Method for Developing Modular Architectures using Genetic Algorithms. *Research in Engineering Design*, 18, 91–109.
- Zotteri, G., Kalchschmidt, M., & Caniato, F. (2005). The Impact of Aggregation Level on Forecasting Performance. *International Journal of Production Economics*, 93-94, 479–491.

**Contact: Franck Marle**, Ecole Centrale Paris, Laboratoire Genie Industriel, Grande voie des vignes, 92295 Chatenay-Malabry, France, (33)141131568, franck.marle@ecp.fr