



HAL
open science

Interactions-based risk clustering methodologies and algorithms for complex project management

F. Marle, Ludovic-Alexandre Vidal, Jean-Claude Bocquet

► **To cite this version:**

F. Marle, Ludovic-Alexandre Vidal, Jean-Claude Bocquet. Interactions-based risk clustering methodologies and algorithms for complex project management. *International Journal of Production Economics*, 2013. hal-01204816

HAL Id: hal-01204816

<https://hal.science/hal-01204816>

Submitted on 28 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.

**Interactions-based risk clustering methodologies and algorithms
for complex project management**

Franck MARLE¹, Ludovic-Alexandre VIDAL¹, Jean-Claude BOCQUET¹

Laboratoire Genie Industriel, Ecole Centrale Paris, 92290 Chatenay-Malabry, FRANCE

ABSTRACT

Project risks have never been so present. First, projects are dealing with bigger stakes and facing stronger constraints. Moreover, projects must cope with an ever-growing complexity. Risks have then increased in number and criticality. Lists of identified project risks thus need to be decomposed, for smaller clusters are more manageable. Existing techniques are mainly mono-criteria, based on risks parameters such as nature or criticality value. Limits have appeared since project risk interactions are not properly considered. Project interdependent risks are indeed often managed as if they were independent. We thus propose an interactions-based clustering methodology with associated tools and algorithms. Our objective is to group risks, so that the interaction rate is maximal inside clusters and minimal outside. The final objective is to facilitate the coordination of complex projects by reducing interfaces when dealing with risks. We first model project risk interactions through binary matrix and numerical matrix representation. Then we define our objective function. A linear programming algorithm and two approximate iterative ones are then presented. Possible refinement through the concept of interactions similarity is also proposed. A case study in the entertainment industry is finally presented, providing us information and points of comparison for global conclusions and perspectives.

1 INTRODUCTION

A project is a temporary and unique endeavor undertaken to deliver a result, which generally corresponds to the creation of a unique product or service which brings about beneficial change or added value (PMI, 2004). A new organization within the firm is then needed to perform a project: new processes which must answer project finalities and objectives in terms of values creation must be set up. These new processes are performed thanks to resources (notably project actors) which belong to the created project organizational system. A project is in essence unique, which means that the project organizational system is to be conceived

for each project within a firm (as it is specific to a project). Project organizations are thus in essence temporary organizations. They coexist with permanent organizations which exist within the firm. This coexistence (involving interfaces and dependencies) makes project and project management all the more complex. Moreover, the conception of the project organizational system follows the steps of project phases' identification and analysis, planning and monitoring.

Therefore, when thinking at projects in terms of systems following several phases, many dependencies and interdependencies between phases, sub-systems and other entities can be identified. This can lead to communication and coordination issues when facing decision-making situations. Namely, project complexity, described notably in (Baccarini, 1996), (Edmonds, 1999), (Laurikkala and al., 2001) involves issues in decision-making under complex situations (Phelan, 1995), (Earl and al., 2001), (Vidal and al., 2010). Then, as well as the uncertainty and instability which are inherent to projects, complexity appears to be one of the main risk drivers. Moreover, complexity reduces the awareness of decision-makers and thus the efficiency of their decisions.

As a consequence, this paper proposes an innovative method and its associated tool to assist project risk management under complex contexts by focusing on project risk interdependencies. Our research objective is to group risks into clusters in order to catch inside of them most of project interactions. This is notably to facilitate the coordination among actors involved in the project risk management process.

2 THE NEED FOR A BETTER CONSIDERATION OF PROJECT RISK INTERACTIONS

Project systems are in essence risky, as they are unique, constrained, subject to uncertainty and to complexity. They are composed of many interrelated objects of different natures and must reach many objectives which may be interdependent or even contradictory. This

involves two things in terms of risk management. First, there are many risks of different natures, and it is even impossible to be completely exhaustive when identifying them. It is then mandatory to group risks into smaller and thus more manageable groups, which is detailed in 2.1. Second, these risks are interrelated, meaning that they are not independent events. This causes some issues in decision-making in project risk management, when decisions are to be made about the prioritization of risks and the risk mitigation actions. These issues are not well addressed by traditional approaches as shown in 2.2. Paragraph 2.3 sets the resulting problem and methodology of this paper.

2.1 Classifying project risks by nature and/or by value

Project risk management is classically decomposed into four successive major steps: risk identification, risk analysis, risk response planning and risk monitoring (PMI, 2004).

Risk identification is the process of determining events which, may they occur, could impact positively or negatively project objectives. Risk identification methods are classified according two different families: direct or indirect risk identification (Raz and Hillson, 2005). The number of risks in the generated list may vary from some tens to some hundreds. It is then mandatory to decompose it into subgroups in order to have more manageable items. This list is a priori (included in the methodology) or a posteriori classified according to the nature of the risks (financial, human, technical, schedule, etc...). This process is called clustering by nature.

During risk analysis, risks are prioritized, essentially according to their probability and impact. Risk evaluation scales are often defined in terms of criticality, which is generally a function of probability and impact. The main output of risk analysis is a list or graph, which enables decision-makers to categorize risks as high, medium or low in terms of the chosen indicator (criticality whatever its formulation, or other indexes). This is another kind of clustering, called by value.

Next steps are risk response planning and monitoring. We argue that these steps should be performed after an innovative project risk analysis based on risk interactions since current methods have shown their limits.

2.2 Limits of traditional approaches

The initial goal of risk clustering processes is to facilitate the management of risks in terms of decision-making. In the case of clustering by nature, the main objective is to facilitate the identification of risk owners, whether individual or entities (due to their skills and competencies), and then to facilitate the allocation of additional or contingency resources. In the case of clustering by value, the main objective is to prioritize risks in order to make decisions about the future mitigation plan and the resulting resource allocation. As they are classifying risks based on one of their characteristics, those methods do not include their possible interactions. Therefore, in both cases, there are lots of interactions between clusters. This can result in a lack of coordination between actors when making decisions, due to the lack of awareness of the global impact of one's decision, mainly outside the considered cluster.

Actually, whatever the criteria used for the decomposition of an initial risk list, and whatever the rigor and detail level used, there will always be interactions between risks which do not belong to the same cluster (Marle, 2002). This can notably be underlined when looking at projects through systems thinking (Simon, 1981), (Le Moigne, 1990).

Project management current techniques include classical principles underpinning scientific management: the fragmentation of work and the maximization of visibility and accountability. We can argue that today projects are generally managed with single-link trees (Work Breakdown Structure, PERT, Organizational Breakdown Structure, risk lists) and not as networks (Vidal and al., 2009). In the case of risk management, most of the methods use lists, screening or sorting risks, as seen before. The problem with current methodologies is that

project risk interactions are not clearly included, e.g. in Figure 1, where some links are existing though not managed (dotted lines). Risks are indeed interrelated with complex links.

Insert Figure 1

A previous study we had conducted about 23 risk analysis methodologies enabled to identify complexity-related issues (Marle, 2008). For instance, there may be propagation from one « upstream » risk to numerous « downstream » risks, the climax of this phenomenon being the famous dangers of the domino effect. Traditional methodologies are mainly single-risk oriented, analyzing their multiple causes and multiple consequences. However, some works have been done to model more complex interdependencies between risks. Bayesian Networks for instance link several risks, from multiple inputs to multiple outputs, but they have specific validity conditions: links must be oriented, and they are more adapted to acyclic networks. However, loops are a great danger during projects and are all the more complicated to understand since risks which exist within a loop are likely to be heterogeneous (of different natures). It is possible to tackle this issue using dynamic Bayesian Networks but the effort required to gather conditional probabilities and to run dynamic networks for the huge number of possible loops makes them unsuitable to real project environments.

There is thus crucial need for better awareness, consideration and management of project risks, knowing they are intertwined. We propose in this article such a methodology. Our ambition is not to give “exact” results: we want to assist day-do-day project risk management thanks to our method. This one is notably not based on the mathematics of probabilities. It can thus take into account easily the existence of loops and non-linear couplings for instance.

2.3 Problem setting and methodology

As shown by the former paragraphs, risks are managed thanks to the elaboration of smaller clusters. At this stage, a management issue arises, since decisions may be blocked, slowed down or ineffective if interactions are poorly taken into account. Our research problematic is thus to propose a new additional clustering methodology, which could take into account interactions between risks, in terms of existence and strength. First, we identify possible risk interactions. The whole is synthesized thanks to binary matrix representation. The matrix is then transformed to be a numerical one thanks to the use of the principles of the AHP (Analytic Hierarchy Process, a multi-criteria decision-making methodology) (Saaty 1980). These numerical data permit us to develop a linear programming and two approximate iterative algorithms. We then express how these results can still be refined thanks to the introduction of a distance measure and similarity identification process. All the obtained results are then compared to classical decompositions, notably thanks to a case study in the entertainment industry. We then propose some conclusions and call for future perspectives of research around this issue, after studying the implications of our works on day-to-day management.

3 CATCHING PROJECT RISKS INTERACTIONS THROUGH MATRIX REPRESENTATION

3.1 Building up the Risk Structure Matrix (RSM)

3.1.1. The Design Structure Matrix approach

The Design Structure Matrix (DSM) represents relations and dependencies among objects. The same objects are both in the rows and columns of the square matrix, which is square. The DSM was introduced by Steward (Steward, 1981) with tasks and was initially used basically for planning issues (Eppinger and al., 1994). Since, it has been widely used with other objects, like product components, projects or people (Eppinger and Salminen, 2001), (Sosa and al.,

2004), (Danilovic and Browning, 2007), (Sosa, 2008). As for us, we propose to use the concept of DSM for other conceptual objects, which are risks, in the context of project management. As tasks, projects and people, project risks are (or can at least be supposed as):

- in a finite number (since a project is in essence temporary, with finite resources, objectives, means, etc., i.e. a finite number of elements),
- managed during the project management process,
- interrelated, (notably because of project and project management complexity factors (Vidal and Marle, 2008)) which justifies the use of a methodology for complex interactions management.

We define risk interaction in terms of the existence of a possible precedence relationship between two risks R_i and R_j . We then define the binary Risk Structure Matrix (RSM). It corresponds to the square matrix with $RSM_{ij}=1$ (else 0) when there is an interaction from R_j to R_i . The main advantage of this approach is to overcome the display issue of complex network and to permit easier calculations which are inherent to the matrix format (eigenvalues, matrices product, matrix transposition, etc...). The combination of several individual interactions may form loops in the risk network (for instance, if $RSM_{ij}=RSM_{jk}=RSM_{ki}=1$, then we have a loop including the three risks R_i , R_j and R_k).

3.1.2. Building up the RSM

In order to build the RSM, we have to identify the interactions which exist between project risks. The iterative procedure we use is notably addressed in ongoing publications. Classically, the DSM is re-ordered in a way which permits to show first-level blocks, thanks to the well-established partitioning process (Eppinger and al., 1994). This one applied to the RSM gives three types of information:

- the dependent risks: they are engaged in a potential precedence relationship,
- the interdependent risks: they are engaged in mutually dependent relation, directly or

with a bigger loop,

- the independent risks: the risks are basically non-related.

The aim of this process is basically to obtain a lower block-triangular matrix. Partitioning enables to isolate interdependent risks, but our purpose is different. We aim at grouping risks in clusters with maximal internal interactions and minimal inter-clusters interactions.

In order to do so, we firstly use an AHP-based evaluation to transform the RSM into a numerical matrix which is to catch the strength of local interactions. Indeed, fieldwork proves us that such assessment of interactions is hard to do directly. On the contrary, it can be observed that people find it easier to say that a cause C_1 is more likely to produce an effect E (first level neighbor) than another cause C_2 , or similarly, that an effect E_1 is more likely to be the consequence of a cause C than another effect E_2 . That is why we claim for the use of the principle of pair wise comparisons to assess project risk interactions.

3.2 Methodology

Five steps are necessary to carry out our methodology. Figure 2 shows this process with an example for risk R_4 :

Step 1: Decomposing individual sub-problems

The presence of a 1 in the binary RSM expresses the existence of a possible precedence relationship between risks R_i and R_j . $RSM_{ij}=1$ implies two possible ways to address the situation. If there is a cause-effect relationship between R_i and R_j , then it is equivalent to consider R_i as a cause of R_j or R_j as an effect of R_i . Similarly as in (Chen & Lin, 2003) for design tasks, we combine these visions. Two stages must thus be performed. For each R_i , we isolate the risks which are related with R_i in column (possible effects) and in row (possible causes). This permits a sanity check because each relationship has to be expressed two times. This identification enables to generate the Binary Cause or Effect Vectors, which are relative to one risk R_i (respectively $BCV|R_i$ and $BEV|R_i$).

Step 2: Evaluating the strength of interactions

We build up two matrices regarding the risk R_i based on the two previously isolated sets of risks (in rows and in columns), which constitute the set of alternatives. They are called Cause or Effect Comparison Matrices and are both relative to one risk R_i ($CCM|R_i$ and $ECM|R_i$).

Given the AHP numerous applications in the field of project management and project risk management (Gourc, 2006), we claim for the use of the AHP-based principle of pair wise comparisons to assess project risk interactions (as we define them in this article). The Analytic Hierarchy Process was developed by Thomas Saaty (Saaty, 1977), (Saaty, 1980). It is a multi-criteria decision-making method. It permits the relative assessment and prioritization of alternatives. The AHP is based on the use of pair wise comparisons, which lead to the elaboration of a ratio scale. The AHP models the problem as a hierarchy, consisting of an overall goal, a group of alternatives, and a group of criteria which link the alternatives to the goal. Pair wise comparisons are carried out by asking how more valuable an alternative A is to criterion C than another alternative B. Pair wise comparisons finally constitute square matrices, the values of which are between 1/9 and 9, and the diagonal elements of which are equal to 1 while the other elements verify two conditions:

- The i - j^{th} element is equal to the comparison between i and j regarding the considered criterion.
- For i different from j , the i - j^{th} element is equal to the inverse of the j - i^{th} element.

This piece of information is processed mathematically, in order to transform the qualitative user information into a mathematical one. Priorities are then determined thanks to these matrices and a global consistency test can be performed to evaluate the coherence of the judgments. The final result is a table which gives a global evaluation of each alternative for the objective and for each criterion.

In our case, we have two parallel pair wise comparison processes to run. The first one consists in the ranking in rows for each project risk. Given the risk R_k , the set of alternatives are all the

non-zero elements of risks other than the diagonal element in row k . The criterion on which the alternatives are evaluated is the contribution to R_k in terms of risk input: in other terms, for every pair of risks which are compared, R_i and R_j (thus following $RSM_{ki}=RSM_{kj}=1$), the user should assess which one is more important to risk R_k in terms of probability to be a risk input (i.e., a cause) for risk R_k . Numerical values express these assessments thanks to the use of the traditional AHP scales. The second one is the ranking in columns, according to the same principles.

Step 3: Consolidating the results

Eigenvectors of each matrix $ECM|R_i$ and $CCM|R_i$ are now to be calculated. It enables to find the principal eigenvectors, relative to the maximal eigenvalue. They are called the Numerical Cause or Effect Vectors and are relative to one risk R_i (NCV_i and NEV_i). Consistency of the results should be tested thanks to the AHP consistency index.

Step 4: Aggregating the results

For each risk R_i , Numerical Cause or Effect vectors are respectively aggregated into Numerical Cause / Effect Matrices (NCM and NEM). The i -th row of NEM corresponds to the eigenvector of $CCM|R_i$, which is associated to its maximum eigenvalue. The j -th column of NCM corresponds to the eigenvector of $ECM|R_j$, which is associated to its maximum eigenvalue.

Step 5: Compiling the results

The two previous matrices are aggregated into a single Risk Numerical Matrix (RNM), the values of which assess the relative strength of local interactions. The RNM is defined by a geometrical weighting operation (based on the possible assumption that both estimations can be considered as equivalent). We choose the geometrical mean instead of the arithmetic mean because it tends to favor balanced values (between the two assessments).

$$RNM (i, j) = \sqrt{NCM (i, j) \times NEM (i, j)}$$

$$\forall (i, j), 0 \leq RNM(i, j) \leq 1$$

The RNM thus permits to synthesize the existence and strength of local precedence relationships between risks as it combines the cause-oriented vision and the consequence-oriented vision of an interaction. Combining these two visions is therefore interesting to avoid any bias or misevaluation which can happen when looking at the problem with only one vision.

Similarly as when performing risk identification and analysis, the identification and evaluation of interactions should be updated during the project.

Insert Figure 2

4 PROPOSED CLUSTERING ALGORITHMS AND REFINEMENTS

4.1 Overall problem definition

We want to cluster risks in order to maximize intra-cluster interactions thanks to the use of the RNM. We do insist on the fact that the values of the RNM are local judgements, which implies that risk interactions assessments are in essence relative. However, we do argue that this first clustering is useful, since it permits to focus on the most significant local risk interactions. In order to define our clustering approach, we carried out a literature review on clustering and graph partitioning approaches, notably thanks to the works of Schaeffer (Schaeffer, 2007). There are mainly two approaches for clustering operations: vertex similarity-based methodologies and cluster fitness measures-based methodologies.

4.1.1. Vertex similarity-based criteria and methodologies

On one hand, there are several clustering algorithms which are based on similarities between the vertices. These methods are based on the assumption that the higher the vertex similarity,

the stronger the need to cluster the vertices together. These measures are mainly based on additional properties of vertices which permit to compute a similarity matrix.

Rather than defining similarity measures, dissimilarity measures such as distance measures are usually defined, for instance the traditional Euclidean and Manhattan distances (Ben-Arieh and Sreenivasan, 1999), (Hennig and Hausdorf, 2006). More advanced distance such as Jaccard distance (Dong and al., 2006) or the Levenshtein distance (Gusfield, 1997) can be used to answer this issue. Some other coefficients can also be calculated to evaluate vertex similarity and perform the corresponding clustering process: for instance there exists angle measures such as the cosine similarity (Lakroum and al., 2005). Some important works even try to compare such similarity measures and their impact on clustering operations (Hartigan, 1975), (Yin and Yasuda, 2006). As noticed by (Schaeffer, 2007), “in some applications, the vertices lack additional properties and there is nothing in the vertices themselves that would allow the computation of a similarity matrix”. In this case, vertex similarity measures are often defined thanks to the structural characteristics of the graph. For instance, some measures based on the correlation of the adjacency matrix can be used, such as the Pearson correlation (Rodgers and Nicewander, 1988) or the Mahalanobis distance (Mahalanobis, 1936).

4.1.2. Cluster fitness measures-based criteria and methodologies

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 fulfil a certain property. For instance, methodologies based on graph density measures have been developed in order to partition the initial graph into subgraphs, the density of which should be inferior and/or superior to chosen values (Karp, 1977), (Kim, 2003), (Zotteri and al., 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, cutsize-based measures undoubtedly permit to quantify the relative independence of a subgraph to the rest of the graph and have been used in many clustering processes (Shi and Malik, 2000), (Kannan and al., 2004). Finding the partition which minimises cut-sizes (with restriction conditions on the orders of the subgraphs) permits to maximise the sum of the edges weights which are internal to the clusters. This cut-based measure seems very interesting in our case. In order to facilitate complex project risk management, one is likely to want to reduce interfaces in terms of number, and above all strength. Reducing interfaces is thus very similar to this problem of graph partitioning which aims at minimising the global cut size (since risk interactions are modelled and assessed thanks to edges and their weights).

4.1.3. Broad definition of the interactions-based project risk clustering problem

Indeed, let us consider a set of project risks (R_1, R_2, \dots, R_N). This set of risks is in essence a complex one, since interactions do exist between risks. Let us suppose we know the RNM of this set of risks (the former steps to build the RNM should have been followed by the user). Let K be the number of clusters of the optimal clustering solution, which maximises intra-cluster global interactions value. INTRA value is defined by the sum of the values of all interactions between risks which belong to a same cluster. INTER (Inter-cluster global interactions) value is defined by the sum of the values of all interactions between risks which are not paired inside a same cluster. The sum of INTRA and INTER values corresponds to the sum of all risk interactions values, which is constant. As a consequence, maximizing INTRA is equivalent to minimizing INTER.

K is not known in advance, but we know some constraints about it. Namely, the goal is to assign project members to clusters in order to manage more properly the risks which belong to a same cluster, i.e. which are strongly interdependent. It is known that people have a

limited capacity to manage simultaneously numerous objects. Indeed, the number of factors that most decision-makers can handle simultaneously cannot be higher than 7 to 9 (Miller 1956). We follow this hypothesis and in the end the maximum size of a cluster should be 9.

We choose to leave some margin compared to the classical empirical rule of 7 objects to be managed simultaneously. This permits us to know a lower bound of K , which is

$K_{\min} = INT \left(\frac{N-1}{9} \right) + 1$, where INT is the integer part of a real number. This problem

corresponds to the K -graph partitioning / clustering problem (Schaeffer, 2007) and we define it more precisely in the following of this article.

4.2 Formulating a linear programming problem to answer this question

Here is the corresponding quadratic integer programming problem formulation. This problem is to be solved for each value of K which is superior to K_{\min} . We first introduce the following decision variables:

$\forall i, 1 \leq i \leq N, \forall k, 1 \leq k \leq K, x_{ik} = 1$ if risk R_i belongs to cluster C_k .

The objective function, which is to be maximized, is as following in equation 2

$$INTRA = \sum_{k=1}^K \sum_{i=1}^N \sum_{j=1}^N x_{ik} x_{jk} RNM(i, j) \quad (2)$$

Problem constraints are the following (equations 3 and 4).

$$\forall i, 1 \leq i \leq N, \sum_{k=1}^K x_{ik} = 1 \quad (3)$$

as we argue for clusters disjunction in order to permit easier management in practice.

$$\forall k, 1 \leq k \leq K, \sum_{i=1}^N x_{ik} \leq 9 \quad (4)$$

since we want the maximum size of clusters to be 9 risks.

This problem is not linear but we can make it easily linear thanks to the introduction of new decision variables (equation 5) and new constraints (equation 6).

$$\forall i, 1 \leq i \leq N, \forall j, 1 \leq j \leq N, \forall k, 1 \leq k \leq K, y_{ijk} \text{ is a binary variable} \quad (5)$$

We define y_{ijk} by adding the constraints:

$$\forall i, 1 \leq i \leq N, \forall j, 1 \leq j \leq N, \forall k, 1 \leq k \leq K, y_{ijk} \leq x_{ik} + x_{jk} - 1 \quad (6)$$

This equation forces y_{ijk} to be equal to 0 if x_{ik} and x_{jk} are not both equal to 1, i.e. if R_i and R_j do not belong to the same cluster. All other constraints are kept for problem formulation. The objective function can then be re-written thanks to these new decision variables, in equation 7.

$$INTRA = \sum_{k=1}^K \sum_{i=1}^N \sum_{j=1}^N y_{ijk} RNM(i, j) \quad (7)$$

We use OPL (Optimization Programming Language) to solve this problem. However, its complexity is high (2^{N-1}), and problems over 20-21 risks appear to be critical when testing them. We presently work on developing some heuristics to assist problem solving in the software and reduce processing time by avoiding trivial non-solutions. Meanwhile, we have developed some approximate iterative algorithms, which permit us to approximate the optimal solution of the problem.

4.3 Using simple heuristics to solve this problem

Both of these heuristics are iterative construction algorithms, but they use two different values for clustering conditions, as described in equations 8 and 9. The first iterative algorithm IA_1 is based on the maximum value between two separate clusters. The second one IA_2 is based on global interactions value between two separate clusters. In the two cases, these values are to be maximized at each step.

$$Value_1(C_\alpha, C_\beta) = \max_{i \in C_\alpha, j \in C_\beta} RNM(i, j) \quad (8)$$

$$Value_2(C_\alpha, C_\beta) = \sum_{i \in C_\alpha, j \in C_\beta} RNM(i, j) + RNM(j, i) \quad (9)$$

At the initial step, all risks are isolated: every initial cluster is a singleton. The maximum value is obtained for two isolated risks R_{i0} and R_{j0} , which are grouped into a first cluster C_1 . Then, the maximum value for the second step may be obtained for two isolated risks or for cluster C_1 and an isolated risk: in the first case, this forms another cluster C_2 and in the second case, C_1 is updated with the third element. Step by step, clusters are formed with singletons or by expanding (adding one singleton to an existing cluster) or by merging (two existing clusters). At each step, the previously defined value ($Value_1$ or $Value_2$) is maximized. In the case the maximum size of a cluster is reached, the second maximum value is identified and the clustering operation is done on the corresponding interaction if it does respect the formulated constraints. This procedure is repeated iteratively until reaching a solution which respects all the constraints.

4.4 Developing indicators to compare solutions

In order to compare different possible clustering solutions, we introduce some numerical indicators. Given a problem with N risks, the first two indicators we introduce are:

- MPT = Mean Processing Time, which is the mean time to obtain the solution of the problem.

- $\nabla INTRA(A, B) = \frac{INTRA(A) - INTRA(B)}{INTRA(A)}$

(10)

where A and B are two solutions and $INTRA$ is the value of the intra-cluster global interactions value. These indicators permit to assess the overall performance of each method (in terms of time and capture of interactions inside clusters).

Given a possible solution thanks to a clustering method, we can build a $K \times N$ matrix (M_{ki}) , so that $M_{ki}=1$ if risk R_i belongs to cluster C_k in the final solution. If ${}^T(M_{ki})$ is the transpose of this matrix, then:

- $H=(M_{ki}).{}^T(M_{ki})$ is a $K \times K$ matrix, the diagonal terms of which correspond to the number of risks which are clustered in cluster K . $MCS = \text{Mean Cluster Size}$ is then the average of the diagonal values of this matrix.
- $L={}^T(M_{ki}).(M_{ki})$ is a $N \times N$ matrix, the i - j^{th} term of which is equal to 1 if R_i and R_j belong to a same cluster. As a consequence, the calculation of $\nabla M(A, B)$ which is the difference of the two matrices $L(A)$ and $L(B)$, obtained thanks to the clustering methods A and B , permits to identify the similarity between two clustering solutions. Let N_0 be the number of non-zero values in the $\nabla M(A, B)$ matrix. We build up the following indicator (Eq. 11) which is a dissimilarity measure when comparing solutions. Note that MCS is taken into account, since, if given a clustering solution, if one risk R_i is taken out of the cluster it belongs to, then MCS non-zero values are likely to appear (mean value) in the dissimilarity matrix for this risk R_i .

$$\nabla(A, B) = \frac{N_0}{N^2 \times \left(\frac{MCS(A) + MCS(B)}{2} \right)} \quad (11)$$

These other indicators permit to compare the structure of two solutions proposed by different methods.

4.5 Refinement of the obtained solutions

Our goal here is to refine our results by identifying within clusters similar situations in terms of causes and effects, i.e. the less distant risks. Many distances (i.e. similarity functions) can be proposed in order to assess the proximity of two risks as seen in 4.1.1. To define them, we build up a symmetrical matrix, the Risk Interaction Matrix (RIM), the i - j^{th} term of which is given by equation 12 :

$$RIM(i, j) = \frac{RNM(i, j) + RNM(j, i)}{2} \quad (12)$$

At this stage, note that any metric can be used in order to define this distance. As for us, we claim for the use of the Mahalanobis distance (Mahalanobis, 1936), which corresponds to a weighted Euclidian distance (the weights being determined by the covariance matrix Σ), as shown in equation 13.

$$D_M(i, j) = \sqrt{(V_i - V_j) * \Sigma^{-1} * {}^T(V_i - V_j)} \quad (13)$$

Where V_i is the i -th column of RIM.

We then use a classical average-linkage clustering algorithm (Murtagh, 1983) to identify similar situations inside clusters. Such identifications of similar risks (in terms of interactions) within the clusters obtained in 4.2 and 4.3 permit to give information to the person in charge of the management of the cluster: indeed, two similar risks may be handled with similar managing techniques and/or even the same preventive/curative actions.

5 CASE STUDY

5.1 Introduction of the case study

In order to test the proposals and results of our theoretical study, we carried out a case study in the entertainment industry: the chosen project is the production of a family stage musical in Paris. The project notably encompasses stage, costume, lightning and sound design, casting management, rehearsal management, fund raising and overall project management support activities, etc... Project duration is 6 months before the 9 months staging. Overall budget is around 60000€ without salaries. The project team is composed of 18 people. The case study we present here is based on fieldwork and interviews which were conducted with 1 cast member, 2 creatives, and 1 production team member. We introduce below the initial list of risks in table 1.

Insert Table 1

5.2 Results of the case study

With the previously introduced data, we built up the RNM. In order to do so, a preliminary interview (which last about 1 hour and a half) was realised with the production team member so he could perform the required pairwise comparisons. Before a meeting with all the participants was organised, the study was presented to them and preliminary qualitative interviews were conducted with them in order to have a general idea of the points which would not make consensus. During the meeting, the evaluations of the production team member were exposed and discussed by all the participants. Some changes occurred while discussing and consensus was reached after the meeting, which last about 3 hours. As a whole, 3 weeks were needed to build up the RNM.

We then processed the algorithms which were presented previously. Their results can be seen next page in Figure 3. They only represent two graphs, since in this case the second iterative algorithm (IA2) gave the same result as the linear programming (LP) algorithm. In this figure, they are compared to the two classical clustering results. A general observation is that there are less arcs (risk interactions) outside the clusters in IA1 or LP than in the traditional approaches. Moreover, it appears that the arcs which remain outside the clusters in IA1 or LP tend to be weighted with low values (low strength of risk interactions) compared to the mean value of interactions.

Insert Figure 3

Indeed, as shown afterwards in Figure 4 (synthetic indicators), it must be noted that interactions-based clustering give here much more efficient results in terms of interactions values within clusters, as expressed by the values of ∇^{INTRA} . The linear programming and

iterative algorithms we developed give very interesting results and perspectives for project risk management since in all cases, more than 70% of the interactions values are kept inside the obtained risk clusters (nearly 5 times best than by nature, and 2 times best than by values).

Insert Figure 4

Practical implications on project management are the following ones. Since the formed clusters are more independent (due to a lower level of interactions between clusters), it seems interesting to propose a complementary organisation. For a given cluster C, the nomination of a risk cluster owner RO(C) seems interesting. This person would then assure the global coordination of the management of the risks which are inside cluster C. Since within cluster C, risks tend to be strongly linked, this person has more easily a global view of the possible implications (propagation effects) of any decisions made for a risk which is inside cluster C. In this case, the process to nominate the risk cluster owner was discussed: the participants proposed to organise a meeting for each cluster C with the initial risk owners which are present in cluster C and ask them to nominate the risk cluster owner.

In this specific case, one of the decisions was notably to make one cast member (a female performer) the prime interlocutor with the production team members. This corresponded to assign her to the cluster which grouped notably R₁₀ (low team communication) and R₁₇ (low team motivation). She was then responsible for the identification of triggering events of risks within this cluster and its communication to the production team. She was also a relay between the production team and the cast to avoid miscommunication. This organisation permitted to create a better ambience and motivation than in former projects which had been carried out by the firm. Discontent was notably detected earlier, before it implied a decrease of motivation and thus a decrease of quality for the show. In other terms, instead of focusing

on the project results, thanks to this approach, some root causes of project risks were better addressed than before.

Moreover, these results were refined. Indeed, the risks R_{11} (Bad scenic, lightning and sound design) and R_{12} (Bad costume design) were analysed as very similar thanks to the Mahalanobis distance-based clustering method we use. The person in charge of the corresponding cluster was informed of this similarity. He then decided to group some management actions of these two risks. For instance, a common monitoring and control process was defined for R_{11} and R_{12} (identical frequency of milestones during the project, simultaneous check of the risks and their consequences,...).

Finally, in this case, IA2 and LP gave exactly the same result. IA1 gave a slightly different result in terms of risks regrouping and intra-cluster interactions value. As a consequence, the issue of algorithm performance compared to MPT is to be addressed, since results do not differ much, whereas MPT can vary of around 275%. In order to answer it more properly and validate even more this overall approach and its practical implications on project management, new tests are to be carried out on several projects (it is presently being tested on a large infrastructure project).

6 IMPLICATIONS ON PROJECT MANAGEMENT

Some works in the literature show that, in the context of decision-making within some specific environments, project managers tend to deny, avoid, ignore and/or delay dealing with risks (Kutsch and Hall, 2005). For all practical purposes, the gap between expected and real risk management implementation is significant. Our methodology permits greater communication on project risks and better confidence in risk management activities thanks to two aspects.

First, the evaluation of risk interactions which is performed when building up the RNM implies a two-step process (looking in terms of causes, and then of effects). Information can thus be checked and refined since one interaction should be listed twice (from cause to effect, and from effect to cause): this checking process permits a better confidence in risk identification and risk interaction identification. Even if theory is sometimes difficult to implement in real projects, we argue that the theoretical background of our models can easily be implemented and understood at a reasonable level. The fact that it relies on expert judgements, mainly qualitative judgements, makes it a user-friendly and easily computable tool. The case study indeed proved us that, even in project contexts where project members are not used to work with tools built from design engineering and industrial engineering theories, the whole approach is globally understood.

Second, clustering risks in order to maximize the global intra-cluster interactions value permits to facilitate the coordination of risk management activities. Namely, it underlines the need for cooperation and transversal communication within the project team, especially within the current structure of the project or the organization. It permits greater communication between people, since it does not seek the identification of ownership, responsibility and/or accountability, but the identification of risk interdependencies. After the clustering process, coordination is made by the person who is assigned to the cluster, but communication has been facilitated before, meaning we have less defensive phenomena.

However, this implies that the assignment of a cluster to a team member should be done carefully, as they need to manage risks which are of different natures. They must have enough background, enough global overview of the different components of the project and enough authority to be able to manage people from different entities with different interests. Such project team members should then be able to facilitate communication and group decision-making. They should also show great adaptability since people may not be used to work together, if they were traditionally managing risks by nature (mainly the same entity). The

question is finally to assign the cluster to one of the owners of one or more risks which are inside this cluster, or to assign it to someone who may have a global view and enough authority on most of the risks which are inside the cluster. In some cases, the best assignment is for the most experimented team member or the one who has the highest hierarchical level. In other cases, the best assignment is for the person who is the owner of most of the risks (in terms of number or global gravity).

As a whole, since our work changes the way risks are grouped together, it changes the way decisions are made. Indeed, with our approach, decisions are made less locally since potential implications and propagation phenomena are handled more easily.

7 CONCLUSIONS AND PERSPECTIVES

In this study, we made a comparison between several possibilities for grouping risks in a project. Our aim is not to criticize the use of classical approaches, because they may have slightly different objectives: we refer to them as points of comparison. Our objective is the improvement of coordination through the better recognition and handling of risks interactions. Our works and case study shows potential significant improvements regarding this specific objective.

They also underline the need for a shift in the way project risk management should be approached. Regrouping risks in clusters which maximize the values of risk interactions inside them appears to be a promising approach to handle project risks. Indeed, such clusters are generally to be assigned to project team members. Each person in charge of a cluster can thus manage risks which are closely related in terms of possible causes or consequences. As a whole, this means that complexity-related possible effects can be caught more easily and as a consequence managed more effectively and efficiently. Project coordination is undoubtedly

facilitated with this approach since interface problems are considerably reduced (for inter-clusters links global value is lowered).

This new approach is thus a complementary one to traditional project risk management techniques. For instance, classifying risks by criticality value is still useful for prioritization and resource allocation, and can be performed within the formed clusters. But the fact that a risk with a low criticality value may involve a chain of more dangerous consequences could change the decision of treating this risk or not. The identification of clusters then enables to treat risks whether individually or collectively. Sometimes, it will be more efficient to treat a risk which could be a cause of the rest of the cluster in order to dramatically reduce the probability of propagation. In other cases, the mitigation actions should impact the complete cluster, because of its global criticality and its very interrelated nature.

Lots of aspects of this work and its results may however be discussed. We identify several of them hereinafter.

- Trying to integrate more project dynamics in this (quite static) methodology, that is to say trying to integrate the time interval which exists between a given cause and each of its effects.
- Consolidating the constitution of the binary interaction matrix. For the moment, we have essentially focused on the triggering effect of the cause and effect relationships between risks, but others do exist (such as changing the gravity of a consequence knowing that a given cause has occurred or not). It would then be particularly interesting to create a typology of possible interactions between risk parameters (probability, gravity, criticality, non detection, influence capacity,...). This could facilitate the evaluation of the strength of interactions in the RNM.
- Exploring the sensitivity of this methodology since the evaluation of risk interactions is at the basis of this approach. As defined presently in our methodology, this evaluation is in essence subjective. Therefore, the use of fuzzy methods such as the

fuzzy AHP needs to be explored in order to reduce the subjectivity of the users' judgements.

- More generally, evaluating with more reliability the relative weights of risk interactions. Formally, the AHP has notably received some criticisms which need to be studied for this issue. Other multi-criteria methods should also be explored (such as MACBETH or ELECTRE) and their results should then be compared to the ones obtained with the AHP.

Future research works are thus to be carried out in order to explore these aspects and develop future implications for project complexity and risk management.

8 REFERENCES

- Baccarini, D., (1996) The concept of project complexity – a review. *International Journal of Project Management*, 14(4): pp. 201-204.
- Ben-Arieh, D. and Sreenivasan, R. (1999) Information analysis in a distributed dynamic group technology method. *International Journal of Production Economics*, 60-61, 427-432
- Chen, S.-J., and Lin, L., (2003) Decomposition of interdependent task group for concurrent engineering. *Computers and industrial engineering*. 44(4), pp.633-650
- Danilovic, M. and Browning, T., (2007) Managing complex product development projects with design structure matrices and domain mapping matrices. *International Journal of Project Management*, 25: pp. 300-314.
- Dong, Y., Zhuang, Y., Chen, K. and Tai, X. (2006) A hierarchical clustering algorithm based on fuzzy graph connectedness *Fuzzy Sets and Systems* 157, pp. 1760–1774.
- Earl C., Johnson J. and Eckert C.M.,(2001) Complexity in planning design processes. *International Conference on Engineering Design. Glasgow, Scotland.*
- Edmonds, B., (1999) Syntactic measures of complexity, in faculty of arts. *Ph.D. Thesis. University of Manchester: Manchester.*
- Eppinger, S., Whitney, D., Smith R. and Gebala D., (1994) A model-based method for organizing tasks in product development. *Research in Engineering Design*, (6): pp. 1-13.
- Eppinger, S. and Salminen, V., (2001) Patterns of product development interactions. *International Conference on Engineering Design. 2001. Glasgow, Scotland.*
- Gourc, D. (2006) - Vers un modèle conceptuel du risque pour le pilotage et la conduite des activités de biens et de services. Propositions pour une conduite des projets et une gestion des risques intégrées. - *Dossier d'habilitation à diriger des recherches présenté à l'Institut National Polytechnique de Toulouse.*
- Gusfield, D. (1997) Algorithms on strings, trees and sequences *Cambridge University Press.*
- Hartigan, J.A., (1975) Clustering Algorithms. *Wiley*
- Hennig, C. and Hausdorf, B. (2006) A robust distance coefficient between distribution areas incorporating geographic distances *Systematic Biology*, 55(1): pp. 170-175
- Kannan, R., Vempala, S. and 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*, vol.2, (no.3), pp.209-24.
- Kim, S. (2003) Graph theoretic sequence clustering algorithms and their applications to genome comparison in Wu, C.H., Wang, P. and Wang, J.T.L. (Eds.): *Chapter 4 in Computational Biology and Genome Informatics*, World Scientific, Singapore
- Kutsch, E. and Hall, M. (2005) Intervening conditions on the management of project risk: Dealing with uncertainty in information technology projects, *International Journal of Project Management*, Volume 23, Issue 8, pp. 591-599.
- Lakroum, S., Devlaminck, V., Terrier, P., Biela-Enberg, P. and Postaire, J.G. (2005) Clustering of the Poincare vectors *Image Processing, 2005. ICIP 2005. IEEE International Conference* Volume 2, Issue , pp. 11-14
- Laurikkala H., Puustinen E., Pajarre E. and Tanskanen K., (2001) Reducing complexity of modelling in large delivery projects. *International Conference on Engineering Design. Glasgow, Scotland.*
- Le Moigne, J.-L., (1990) La théorie du système général. Théorie de la modélisation. *Presses Universitaires de France.*
- Mahalanobis, P C (1936). On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India* 2 (1): pp. 49–55.
- Marle, F., (2002) Modèle d'informations et méthodes pour aider à la prise de décision en management de projets, *Ph.D. Thesis. Ecole Centrale Paris: Paris.*
- Marle, F., (2008) Analysis of current project risk management methodologies and required improvements, Ecole Centrale Paris. Working Paper
- Miller, G. (1956). "The magical number seven plus or minus two: some limitations on our capacity for processing information." *Psychological Review* 1956(63): pp. 81-97
- Murtagh, F., (1983) A survey of recent advances in hierarchical clustering algorithms. *The Computer Journal*. Volume 26, Number 4, pp. 354-359
- Phelan, S., (1995) From chaos to complexity in strategic planning. *55th meeting of Academy of Management. Vancouver*
- PMI, S.C., (2004) A guide to the project management body of knowledge (PMBOK). *Newton Square, PA, USA. : Project Management Institute.*
- Raz, T. and Hillson, D., (2005) A comparative review of risk management standards. *Risk Management: An international journal*, 7 (4): pp. 53-66.
- Rodgers, J. L. and Nicewander, W.A. (1988) Thirteen Ways to Look at the Correlation Coefficient *The American Statistician*, 42, pp. 59-66.
- Saaty, T., (1977) A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, 15, pp. 234-281.
- Saaty, T., (1980) The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation. *McGraw-Hill.*
- Schaeffer, S.E. (2007) Graph clustering *Computer Science Review* I, pp. 27-64
- Shi, J. and Malik, J. (2000) Normalized Cuts and Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v.22 n.8, pp. 888-905
- Simon, H., (1981) The Sciences of the artificial. *Cambridge The MIT Press.*
- Sosa, M., Eppinger S. and Rowles C. (2004) The misalignment of product architecture and organizational structure in complex product development. *Management Science*, 50(12): pp. 1674-1689.
- Sosa, M., (2008) A structured approach to predicting and managing technical interactions in software development. *Research in Engineering Design.*
- Steward, D., (1981) The Design Structure Matrix: a method for managing the design of complex systems. *IEEE Transactions in Engineering Management*, 28(3): pp. 71-74.
- Vidal, L.A. and Marle, F., (2008) Understanding project complexity: implications on project management. *Kybernetes, the International Journal of Systems, Cybernetics and Management Science.*

- Vidal, L.A., Marle, F. and Bocquet, J.C. (2009) - Interactions-based clustering to assist project management. *International Conference on Engineering Design. 2009*. Stanford, United States of America.
- Vidal, L.A., Marle, F. and Bocquet, J.C. (2010) – Measuring project complexity using the Analytic Hierarchy Process. *International Journal of Project Management*. 2010. doi:10.1016/j.ijproman.2010.07.005
- Yin, Y. and Yasuda, K. (2006) Similarity coefficient methods applied to the cell formation problem: A taxonomy and review. *International Journal of Production Economics* 101, pp. 329–352
- Zotteri, G., Kalchschmidt, M. and Caniato, F. (2005) The impact of aggregation level on forecasting performance *International Journal of Production Economics* 93–94, pp. 479–491

Risk	Category	Proba	Impact	...
R1	Human			
R2	Human			
R3	Human			
R4	Technical			
R5	Technical			
R6	Financial			
R7	Financial			
R8	Financial			
R9	Financial			
R10	Schedule			
R11	Schedule			
R12	Schedule			
R13	User			
...	...			

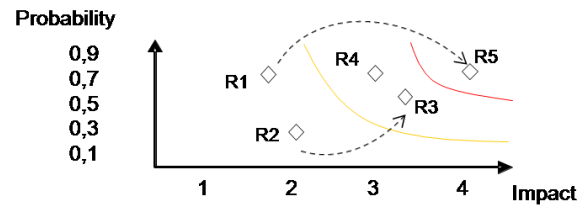


Figure 1. Classification of projects risks by nature and/or by value

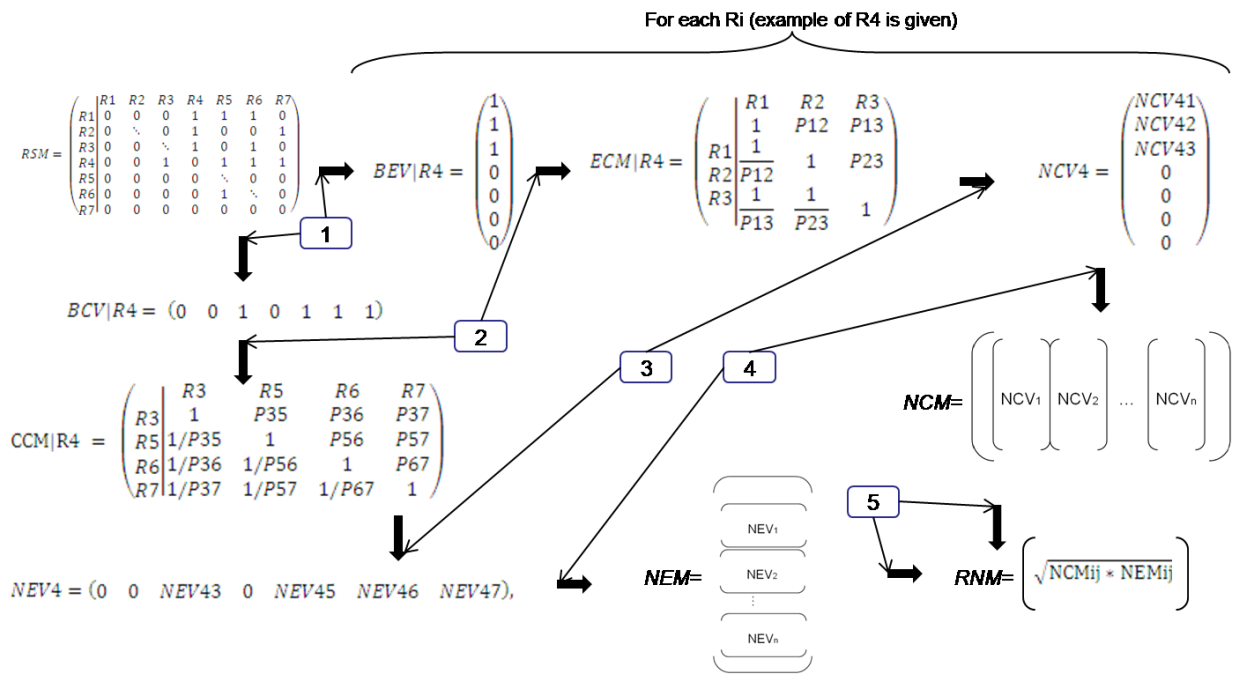


Figure 2. How to build the Risk Numerical Matrix (RNM)

Table 1 : description of the initial list of risks for the entertainment project

ID	Risk Name	Nature	Criticality	Probability	Gravity	P*I
R01	Low budget	Cost & time	Unacceptable	8	7	56
R02	Law and regulations infractions	Contracts	Unacceptable	7	5	35
R03	Low communication and advertising for the show	User / Customer	Unacceptable	8	9	72
R04	Unsuitable cast	Organization	Unacceptable	5	9	45
R05	Unsuitable ticket price setting	Strategy	Unacceptable	7	6	42
R06	Unsuitable rehearsal management	Controlling	Acceptable	3	8	24
R07	Cancellation or delay of the first performance	Cost & time	Unacceptable	5	8	40
R08	Poor reputation	User / Customer	Acceptable	3	7	21
R09	Lack of production teams organisation	Organization	Acceptable	4	6	24
R10	Low team communication	Organization	Acceptable	3	6	18
R11	Bad scenic, lightning and sound design	Technical performance	Negligible	2	7	14
R12	Bad costume design	Technical performance	Acceptable	3	8	24
R13	Low complicity between cast members	Technical performance	Acceptable	3	7	21
R14	Too ambitious artistic demands compared to project means	Requirements	Acceptable	7	2	14
R15	Few spectators / Lukewarm reception of the show	User / Customer	Acceptable	2	9	18
R16	Technical problems during a performance	Technical performance	Acceptable	4	5	20
R17	Low cast motivation	Organization	Negligible	2	4	8
R18	Childish direction (unsuitable for family audiences)	Strategy	Negligible	2	5	10
R19	Low creative team leadership	Controlling	Unacceptable	3	10	30
R20	Low creative team reactivity	Controlling	Negligible	2	2	4

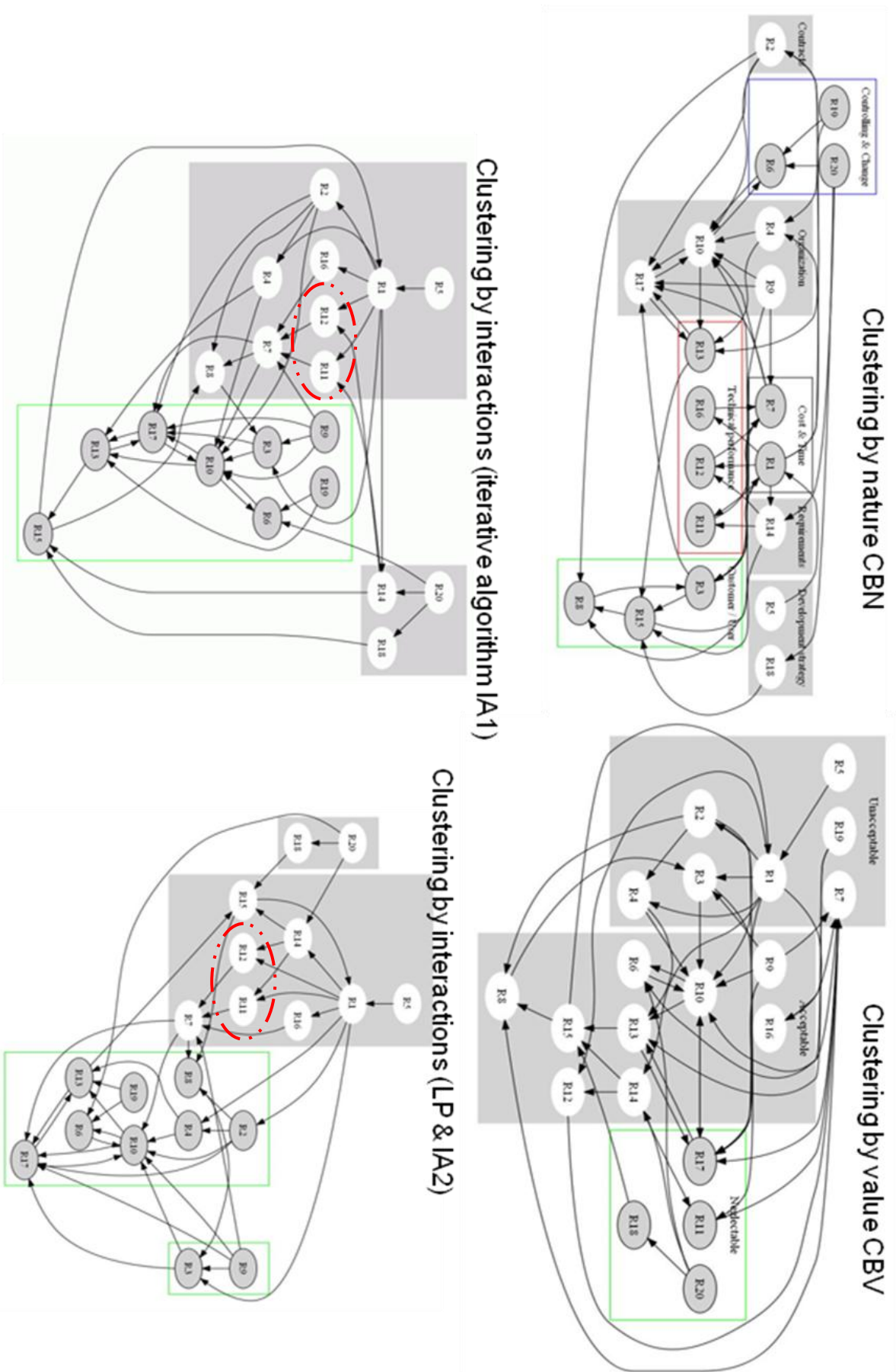


Figure 3. Results of the clustering algorithms

	Intra	MPT	MCS	Intra %
Clustering by nature (CBN)	1,586	0 min	2,5	13,96%
Clustering by value (CBV)	4,361	0 min	6,67	38,39%
Clustering by interactions (LP)	8,376	40 min	5	73,74%
Clustering by interactions (IA 1)	8,247	15 min	6,67	72,60%
Clustering by interactions (IA 2)	8,376	18 min	5	73,74%

$\nabla INTRA(A, B)$

	CBN	CBV	CBI-LP	CBI-IA1	CBI-IA2
Clustering by nature (CBN)	0,00%	-174,97%	-428,12%	-419,99%	-428,12%
Clustering by value (CBV)	63,63%	0,00%	-47,93%	-89,11%	-92,07%
Clustering by interactions (LP)	81,06%	47,93%	0,00%	1,54%	0,00%
Clustering by interactions (IA 1)	80,77%	47,12%	-1,56%	0,00%	-1,56%
Clustering by interactions (IA 2)	81,06%	47,93%	0,00%	1,54%	0,00%

$\nabla M(A, B)$

	CBN	CBV	CBI-LP	CBI-IA1	CBI-IA2
Clustering by nature (CBN)	0	0,355	0,08666667	0,345	0,08666667
Clustering by value (CBV)	0,355	0	0,07204117	0,4	0,07204117
Clustering by interactions (LP)	0,08666667	0,07204117	0	0,03945111	0
Clustering by interactions (IA 1)	0,345	0,4	0,03945111	0	0,03945111
Clustering by interactions (IA 2)	0,08666667	0,07204117	0	0,03945111	0

Interaction Total **11,3589347**

Figure 4. Performance of the clustering methods – Case study