Forming risk clusters in projects to improve coordination between risk owners

Franck Marle, Ludovic-Alexandre Vidal

Forming risk clusters in projects to improve coordination between risk owners.

Franck Marle[*,1] and Ludovic-Alexandre Vidal[2]

1 : AssociateProfessor, Laboratoire Genie Industriel, Ecole Centrale Paris, 92290 Chatenay-Malabry, France.
Email : franck.marle@ecp.fr

2 : Assistant Professor, Laboratoire Genie Industriel, Ecole Centrale Paris, 92290 Chatenay-Malabry, France.
Email :ludovic-alexandre.vidal@ecp.fr

*Corresponding author: franck.marle@ecp.fr

ABSTRACT:

Due to the growing complexity of projects, their risks have increased in number and criticality. Risk lists thus need to be broken down into smaller, more manageable clusters. Classical clustering techniques are generally based on a single parameter, like risk nature, criticality or ownership. Risk interactions are therefore not properly considered when building up clusters. That is why this paper aims at grouping risks so that the communication and coordination between the actors who are committed in the management of the project and its risks are facilitated. Thework is based on an optimization algorithm which maximizes interaction rate within the risk clusters. This paper focuses on two additional points. First, the optimization problem formulation is enriched by some constraintsrelated to the risk owners, not only to the risks. Second, a frequency approach is introduced, to test different configurations, in order to improvethe robustness of the clustering decision. It enables meaningful and operationally realistic actors groups to be organized, regarding not only the interaction rate between risks but also the relationships between risk owners. Our clustering approach encourages people to meet together and communicate/ coordinate better, which we hope will contribute to prevent some undesired complex phenomena.

Keywords : Project risk management; Complexity; Risk interactions; Risk network; Clustering; Coordination; Project organization

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, 2008). As a whole, project management appears to be a complex and risky activity, which underlines the need for efficient and effective project risk management. Projects are in essence complex, due to their size, variety, interdependences and context dependence (Vidal et al., 2010).Project complexity, such as that described in (Baccarini, 1996), (Edmonds, 1999), (Laurikkala et al., 2001), (Earl et al., 2001) involves specific issues in decision-making under complex situations. Indeed, the complexity of a project makes it impossible to have complete information about the project in question and thus to simultaneously visualize all the elements and interactions of a given project. This is underlined when looking at projects through systems thinking (Simon, 1981). In the end, this may lead to failure and dramatic propagation effects because of the interrelated nature of the project elements.Complex phenomena may occur and eventually propagate throughout the project structure. This is likely to reduce the project risk management performance (Eckert et al., 2004) and may have potential consequences on both project processes and results (Kloss-Grote & Moss, 2008).

Project risk management is classically decomposed into four successive major steps: risk identification, risk analysis, risk response planning and risk monitoring (PMI, 2008). 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). This step in the end generates a list of risks. The number of risks in this list may vary from tens to hundreds of risks, managed by more or less owners belonging to more or less different organizations (different companies and/or different departments in a company). It is then mandatory to decompose this list into subgroups in order to have more manageable items. In other terms, project risks need to be clustered.

This paper proposes an innovative method and its associated tool to assist project risk management under complex contexts by focusing on project risk interdependencies. A general approach to clustering project risks is presented. A first version of the optimization problem formulation has been introduced in (Marleand Vidal, 2011). The originality of this paper is to introduce several management-related constraints, such as the maximum number of risk owners within each cluster. This enables additional constraints to be formulated, not only on the risks, but also on the actors who manage the risks. The algorithm proposes a configuration of risk clusters, which is analyzed in terms of parameters related to risks, mainly risk interactions, and in terms of the groups indirectly formed by the actors who own the risks in the clusters.A case study in the field of the construction industry (design and installation of a tramway infrastructure in a city) is finally presented at the end of the paper to illustrate the practical application of these methodologies in fieldwork for large complex projects.

Overview and critique of related works

Literature review about clustering

Clustering is known as the identification of patterns around which communities of elements can be grouped (Gomez et al. 2011). Numerous approaches to cluster elements have been carried out, which may be unsupervised or supervised, and ascending or descending methodologies. To have an extensive overview of clustering methodologies, the authors recommend the reading of (Schaeffer, 2007). Some of the possible approaches are introduced below, considering graph partitioning methods, kernel-based methods and spectral methods.

Most approaches to clustering are supported by a similarity measure between pairs of vertices, commonly defined by a distance, like the classical Euclidian distance, or the Jaccard distance (Dong and al. 2006), or the Pearson correlation in the adjacency matrix (Wasserman and Faust 1994). The partitioning can then be done without knowing k in advance, or requires this information like in the k-means method (McQueen 1967).

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 (Kim 2003). Cut size-based measures permit to quantify the relative independence of a sub graph to the rest of the graph and have been used in many clustering processes (Shi and Malik 2000). Some works 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 (Girvan and Newman 2002, Freeman 1977).

Kernel-based methods are used in cases when classical k-means partitioning algorithms cannot be applied, and are based on the mapping of graph nodes to a higher-dimensional space using a nonlinear function, the kernel (Gomez et al. 2011, Camastra and Verri 2005, Dhillon et al. 2005).

Spectral clustering uses stochastic matrices that denote similarity between connected elements with some uncertainty (Ng et al. 2001). The associated algorithm may simulate information flow, like in the Markov Clustering Algorithm (Van Dongen 2000).

A cluster can contain identical or similar elements, with a particular element called centroid and representative of the group (Filippone et al. 2008). However, in this work, the clustering is made according to the interaction strength between vertices, and is not based on vertex similarity. It is thus formulated mathematically as a K-graph partitioning problem where the objective function is to maximize the interactions value within the clusters, without knowing K in advance. As shown by (Kwak and Anbari 2009), the field of Operations Research/Decision Sciences is very connected to project management research, and we argue that such optimization techniques as the one presented here are valuable for the progress of the project management discipline.

Clustering risks in project management

Indeed, classical ways to cluster project risks are according to their attributes (Marleet al. 2013), respectively their nature or domain or class (called here Clustering by Class), criticality (the classical Probability-Impact product) or ownership(called Clustering by Ownership).The problem with current methodologies is that project risk interactions are not explicitly incorporated.First, they aim at grouping elements according to their similarities (or excluding them of the clusters according to their differences, or distance).Our problem is different, since we are not comparing nodes according to their characteristics, but we are grouping nodes because of the values of the edges that link these nodes. Second, these classical decompositions are based on a single criterion, whether class, criticality or ownership. Grouping risks according to a more sophisticated way could be done by introducing a multi-criteriasimilarity measure, but this is not the object of this work.

There is thus crucial need for better awareness, consideration and management of project risks, knowing they are intertwined. Recent research works have focused on the interactions between project success factors (Chen et al., 2012) to understand better the possible mutual implications of success factors in order to control them better and assist the management of project performance in the case of construction projects. The aim of our paper is to propose a more generic approach, which focuses on risks and permits to clusterthem according to their interaction level. This approach is then all the more interesting than it enables to constitute human groups which are to trigger discussions between project risk managers, the management of which would permit to cope better with possible propagation effects and other undesired complex phenomena. The aim is to adapt the organization to the complexity of potential relationships between risks, knowing that the current official organization is built according to other reasons. This is thus a complementary way to make people communicate and work together and coordinate their decisions. The indirect goal is to assign risk owners to clusters in order to manage more properly the risks which belong to a same cluster, i.e. which are strongly interdependent.

Approaches to thisspecific clustering problem

This problem was firstly introduced in (Vidal et al., 2009). Approaches used to answer this problem are generallybased on the modeling of the network of project risks and their interactions using matrix representations. This can be considered as an extension of the traditional Design Structure Matrix approach since project risks are represented using Risk Matrices, which capture the interactions between risks (Marle and Vidal, 2008). Risk interactions are defined as the possibility for a project risk which has occurred to trigger another one within the project risk network. This risk network modeling has been used in other more recent papers, such as (Allan and Yin, 2011), (Chen et al., 2011) to study other issues. Particularly (Fang et al., 2012)use such models to carry out topological analyses ofproject risk networks and to study the propagation flows within the project risk network and their impact on the traditional evaluation of project risks probabilities and gravities, which has different objectives than clustering.

Our clustering algorithm aims at maximizing the level of interaction among each risk cluster while respecting some constraints related to these clusters and to the human groups derived from risk clusters.Such clustering operation is always feasible since this method does not aim at creating independent (disjoint) risk clusters (which would be impossible in most cases due to the frequent relatively high amount of interactions in complex projects risk networks). Former publications on this issue, notably (Marle and Vidal, 2011), (Marle et al., 2013) addressed this problem, but only through heuristics which could only permit to approximate solutions to the problem. Here, in this paper, we chose to “facilitate” the problem through the introduction of several additional fieldworkmanagerial constraints (maximum number of actors – i.e. risk owners – within clusters, maximum number of groups for each actor, etc.) which permit to have smaller boundaries for the problem.

Unlike the problems presented in former publications, the one presented here can besolved with the C-Plex software and not with heuristics. The constraints on cluster size and number of actors permit to obtain a direct exact solution for problems with less than 60 risks in the project risk network. For larger problems, we recommend the use of some heuristics presented in (Marle and Vidal 2011) to reduce the size of the problem, and then obtain exact solutions for the remaining parts of the problem using C-Plex using the method presented in this paper.

The main originality of this approach is to form human groups considering how risks are clustered. The algorithm aims at maximizing risk interactions within clusters, and proposes risk owner groups corresponding to risk clusters.

Finally, contrary to the formerly cited articles, this paper also introduces later a frequency approach to study the robustness of the results, thus making another improvement of existing project risk clustering methodologies. This frequency approach can also be used to ensure the robustness of the use of the heuristics to reduce the size of large problems.

Formulating the problem

In this paper, the proposed methodology takes into account simultaneously the clusters of risks and the groups of actors who own these risks. These are indirectly formed from the two affiliation relationships, risks to clusters and actors to risks. The following nomenclature is used to formulate the problem.

Nomenclature

NA: the number of actors in the problem

NR: the number of risks in the problem

NC: the number of clusters in the problem

AR: the ownership affiliation matrix of actors to risks.

RR: the risk interaction matrix.

RC: theaffiliation matrix of risks to clusters. It is our decision variable.

AC: the affiliation matrix of actors to clusters, derived from AR and RC.

ClusterSize: the maximum number of risks that each cluster can contain.

ActorSize: the maximum number of actors allowed in each cluster.

MaxGroups: the maximum number of groups that each actor can belong to.

The objective value 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 Equation (1):

$$\max \quad \sum_{0 \le k < NC} \sum_{0 \le j1, j2 < NR} \mathrm{RC}_{j1,k} * \mathrm{RC}_{j2,k} * \mathrm{RR}_{j1,j2} \qquad (1)$$

NR is the number of risks in the problem and **NC**the number of clusters.

RR is a NRxNR matrix with its elements $RR_{j1,j2}(0 \le j_1, j_2 < NR)$ representing the interaction value between the risks j_1 and j_2, already introduced in (Marle and Vidal 2011) as the RNM (Risk Numerical Matrix). This matrix is first built as a binary matrix representing the existence of a potential interaction between couples of risks, then transformed into a numerical one enabling the interaction strength to be assessed. Basically, there are two ways to perform this assessment. The first one is a direct expert evaluation of risk interactions using a Likert scale from 0 to 10, with a possible (not mandatory) normalization of the values in the matrix. But such direct absolute evaluation can be hard to perform even for experts of the project. That is why a second possibility is to have a relative evaluation of risk interactions using pairwise comparisons (stating for instance that interaction 1 is far greater than interaction 2, that interaction 1 is slightly lower than interaction 3, etc.) which can be in the end transformed into numerical values as in (Chen and Lin 2003).

RC Matrix is a NR×NC variable matrix with each of its elements $RC_{j,k}(0 \le j < NR, 0 \le k < NC)$ being a Boolean variable. For each risk, the variable $RC_{j,k}$ being 1 means the presence of Risk *j* in Cluster *k*, while being zero means its absence. RC is our decision variable.

Initial constraints, already introduced in (Marle and Vidal 2011), are related to the inclusion of risks in clusters, and are described by Equations (2) and (3), respectively the maximum number of clusters that a risk can belong to and the maximum number of risks that a cluster can contain:

$$\forall j \in [0..NR-1], \sum_{0 \le k < NC} \mathrm{RC}_{jk} \le 1 \qquad (2)$$

$$\forall\, k \in [0..NC-1], \sum_{0 \leq j < NR} RC_{j,k} \leq ClusterSize_k \quad (3)$$

Where**ClusterSize** is a vector of size NC with its element *Clustersize*$_k$ being the maximum number of risks that the kth cluster can contain.

AR is a NA×NR matrix with its elements ($AR_{i,j,}$ $0 \leq i < NA$, $0 \leq j < NR$) being either 0 or 1, which represents the ownerships of risks for each actor. For example, $AR_{i,j}= 1$ means that Actor *i* is in charge of Risk *j*. This matrix has been generated at the beginning of the project when we did the case study; hence it is not a variable matrix.

AC is a NA×NC variable matrix that has been created to represent the presence of the actors in each cluster, with all its elements being Boolean variables. **AC** is generated from the matrix product of **AR*RC**, which gives the number of times where each actor *i* is present in cluster *k*. **AR*RC** is normalized, in order to get the binary information of the presence of actor *i* in cluster *k*, without considering the number of risks that this actor owns in this cluster. Similar to the **RC** variable matrix, the variable $\mathbf{AC_{i,k}}$being 1 means the presence of Actor *i* in Cluster *k,* while being zero means its absence. This matrix is not a decision variable, it is a consequence of the **RC** variable.

The first additional managerial constraint is to limit the number of actors in the formed groups. Namely, with a cluster of N risks, it is possible to have between 1 and N different actors managing these risks, which is completely different in terms of group management. This is why the ActorSize constraint is introduced, which can be standard or customized by cluster, as formulated in Equation (4):

$$\forall\, k \in [0..NC-1], \sum_{0 \leq i < NA} AC_{i,k} \leq ActorSize_k \quad (4)$$

Where **NA**is thenumber of actors in the problem and **ActorSize** is a vector of size NC with its element *ActorSize*$_k$ being the maximum number of actors in each cluster *k*.

It is also useful to consider the number of groups to which an actor is assigned, in order to avoid potential workload and schedule issues, as described in Equation (5) :

$$\forall\, i \in [0..NA-1], \sum_{0 \leq k < NC} AC_{i,k} \leq MaxGroups_i \quad (5)$$

Where**MaxGroups**is a vector of size NA with its element $MaxGroups_i$ being the maximum number of groups an actor *i* can belong to.

Figure 1 illustrates on a small example the different matrices involved in the process, from the inputs RR and AR to the outputs RC and AC.

Figure 1. Deducing actors groups from risk clusters and risk ownership

The complexity of this problem is due to the mix of constraints which are directly related to the risk clusters and indirectly related to these clusters via the ownership relation between risks and actors. The second issue is that it is difficult for the decision-maker to specify in advance the right configuration of clusters and groups. That is why it is proposed to make these parameters vary, considering an approach based on frequency indicators, described in the following section.

Building up a frequency analysis approach

The approach is based on some variations of some parameters of the optimization problem, in order to compare the proposed solutions, and to count the number of times where risks are put together in a same cluster. The principle of the approach is thus to define the experiments plan to make some parameters of the problem vary, to define some frequency indicators, and then to make decisions knowing the percentage of times when each couples (R_{j1},R_{j2}) are assigned together. In some cases, the possibility that they are assigned to the same cluster is very close to 0% or 100%, they will then be declared respectively as “never” or “always” together. The parameters that may vary are mainly the constraints defined before, the maximum number of clusters for a risk, the maximum number of risks in a cluster, and the maximum number of actors in a cluster.

Frequency indicators

We define N_{Config} as the number of different tested problem configurations. For instance, if we analyze the influence of the uniqueness constraint (included or not, so two possibilities) and of different maximum sizes for human groups (three values, 4 actors, 6 actors and 8 actors), and different maximal cluster sizes (8 or 10), then we get 2*3*2 = 12 configurations.

We introduce in Equation (6) a new index which calculates the percentage of times where two risks are put in the same cluster (Common Cluster Frequency Index). An associated complementary index gives the percentage of times where a risk is included in a cluster (Clustered Frequency Index), introduced in Equation (7). For different configurations $C_l (1 \leq l \leq N_{config})$, we have different results $\mathbf{RC_l}$. The matrix which indicates if two risks are put together in the configuration C_l is called Clustered Organization $\mathbf{CO_l}$.It is the matrix product of $\mathbf{RC_l}$ by its transpose $^{\mathbf{T}}\mathbf{RC_l}$. The global frequency matrix is defined as the sum of all $\mathbf{CO_l}$ for all tested configurations, divided by the number of configurations N_{config}. Non-diagonal terms correspond to the Common Cluster Frequency Index for a couple of risks, and the diagonal terms give the Clustered Frequency Index for arisk :

$$\mathrm{CCFI}(j_1,j_2) = \frac{\sum_{l}^{NConfig} COl(j1,j2)}{NConfig} \tag{6}$$

$$\mathrm{CFI}(j) = \frac{\sum_{l}^{NConfig} COl(j,j)}{NConfig} \tag{7}$$

For each configuration C_l, the matrix CO_l is binary ($\mathbf{CO_{lj1,j2}} = 1$ if and only if risks j_1 and j_2 belong to the same cluster). That means that both indexes are between 0 and 1 (or 0% and 100%). The interesting values are 0% and 100%. CCFI = 0 means that the risks are never clustered together and 100% means that they are always in the same cluster. Similarly, if a risk is always included in a cluster, even if with different risks, then it can give an indication that this risk should preferably appear in the chosen clusters. This can give an indication on the robustness of the decision to put together two risks (if their CCFI = 1), or to keep isolated one risk (if its CFI = 0). It is complementary to the definition of the optimization problems, since it considers the robustness of the decision. The procedure is as following:

1. Step 1 is a screening step for CFI(i) equal to 0. The risks which are never included in a cluster are reordered in the bottom-right part of the matrix.
2. Step 2 is an aggregating step for CCFI(i,j) equal to 1. It gives some clusters, which are or not full and reordered on the top-left part of the matrix.

3. Step 3 is a decision-making process on the middle part of the matrix for inclusion or not of remaining risks in existing clusters.

Several situations may occur at step 3. The closer to 1 the index is, the more the decision is robust to put them together. But, with an index of 70-80%, this is not a safe decision. The worst case is when a risk has an index of 50% within two clusters. It is a kind of dilemma, since half the time this risk has been clustered with the risks of cluster 1 and half the time with risks of another cluster.

Analysis of frequency results

From the analysis of frequency of clustering for risks and couples of risks (CFI_i and $CCFI_{ij}$), it is possible to display the results. It represents the two indicators with a 5-level scale (0, 25%, 50%, 75%, 100%), in order to be easier to read. The rows and columns are reordered in such a way that very dense areas are visible (like kernels), with intermediary areas where percentage is between 25 and 75%. That means that some risks are somewhere between two clusters, and that the decision-maker has to decide whether they are put in one cluster or in the other.

Application

Project description and analysis of the current organization

The industrial background of this study is a large infrastructure project, which consists in building the infrastructure and associated systems of the future tramway of a large city. The lead company is historically a designer/developer of trains, which recently extended its scope by proposing turnkey projects, including the complete infrastructure and equipment around the trains. Risk management has often been mentioned as an important process in the construction industry (Tatum, 1989), (Xue et al., 2010), (Haponava and Al-Jibouri, 2012). In the case of this project, a project risk management process was implemented and led to the identification and assessment of 56 risks managed at the top level of the project. They are classified according to six risk classes (risk nature): contractual, financial, technical, project management, stakeholder management and country. Risk ownership in terms of responsibility is shared by 12 actors in the project. Currently, risk management receives moderate attention within the firm for several reasons (considered too academic, inefficient, money-consuming,…)

A series of interviews were carried out within the organization to identify and assess risk interactions. In the end, the **RR** matrix for the studied risk network was obtained, as shown in Figure 2.

Figure 2. Description of the Risk Interactions Matrix RR for the case study

The affiliation of actors to risks permitted to build the **AR** matrix, displayed in Figure 3.

Figure 3. Risk ownership for the case study (AR matrix)

Due to the number of interactions outside the official project structures, the danger is that some propagation may occur without the organizational capacity to cope with it. When clustering risks according to their nature (one of the traditional approaches in the firm and more generally in project management methodologies), it permits to encompass 44% of interactions within clusters, which is relatively small. More important, when having a look at risks clustered according to their ownership (i.e. the actors who actually manage the risks), only 36% of the interactions are within the groups. This means that if actors do not talk together, 64% of the interactions (and the corresponding propagation effects) might be missed. The aim of the clustering proposed her is then to increase the number of interactions within clusters. A desired consequence is an increase in organizational capacity, and a reduction of potential propagation of the occurrence of one or several risks.

Analysis of clustered organizations

Since we aim at grouping project risks according to their interactions rate, this is inherent to our problem formulation to get heterogeneous clusters. In the end, our clustering approach permits to suggest an organizational structure which is complementary to the existing one(s). The interest of having different structures is to organize meetings with different groups of actors who will exchange on specific aspects of the project (tasks, risks). It is up to the manager to define the number and frequency of group meetings, depending on the complementarities and relevance of each structure.

The reconfiguration of an organization raises the issue of risk ownership and risk cluster ownership.

Indeed, it appears that within clusters, there are numerous different risk owners and often numerous different classes. Interfaces between actors are then highlighted and need to be managed.

The point is to improve coordination between all the risk owners within a same cluster. This reconfiguration may make risk owners more aware of the possible implications of the decisions they make. This is why we decided to test some configurations with constraints on the number of actors in each cluster. This enables the management of the cluster to be facilitated, and in particular the meetings, since the decisions and the communication are sensitive to the number of people inside the group.

There are two possibilities for running the algorithm with the constraint on actors: it is possible to include it in the first run, simultaneously with the other constraints, or to determine it once the first configuration is proposed, since we have a better idea of the "shape" of the clustered organization. In this example, we began by the second type of analysis, it was then impossible to run after that the first type, because it would have been biased. We intend for further works to run the both possibilities in parallel.

With the initial configuration, called CBI (Clustering by Interactions), it appears for some clusters that there are five or six different risk owners for a 9 risk cluster. The algorithm was run again with a constraint of 4 actors at most, called CBI-CA (Clustering by Interactions with Constraints on Actors). The results are shown on figure 4. They are of course lower in terms of optimization of the intra-cluster value, but are still better than the initial non clustered configurations, respectively CBC (Clustering by Class) and CBO (Clustering by Owner), as shown on Table 1.

The obtained clusters seem to be quite consistent with the fieldwork, as they form groups of risks which seem to be relevant in the task of assisting project risk management. Some clusters, for instance, group possible chain reactions which could imply delays (respectively for permits and authorizations, train delivery, depot construction and track installation) and then impact on the final performance indicator which is the profit. The delivery of this part of the project requires simultaneously three things: the depot, the tracks and the trains. If one of these is late, then there is a problem with associated damages. The interesting thing is to mix different risks, for example design-related risks and construction-related risks in the same cluster, in order to show their combined influence on a final issue (for instance the depot with the trains on the tracks). This appears to be all the more interesting since such chain reactions were not previously highlighted and managed during the project.

Figure 4. Proposed organization using Clustering By Interactions taking into account Constraints on Actors (CBI-CA)

When comparing the different clustering alternatives, it can be said that Clustering ByInteractions leads to an important improvement regarding the consideration of interactions. Indeed, the intra-cluster value of CBI is increased by 32% when comparing with CBC and by 61% when comparing with CBO. Moreover, this increase is all the more noticeable given that some risks are left outside clusters in the case of CBI, meaning that the formed clusters are denser. In terms of value, CBI is as balanced as CBO (standard deviation of clusters value) but with a double mean value. Moreover, when adding the constraint on actors (CBI-CA), the results are obviously less optimal than CBI, but still bring substantial improvement compared to traditional approaches. The advantage of CBI-CA compared to CBI is that with this constraint, actors are not too numerous within a discussion group, thus facilitating discussions even more. A corollary is that the standard deviation of number of actors within clusters decreases, thus making more homogeneous groups in terms of size within the organization, which has positive impact on the recognition of work of each actor : people less feel that they belong to "small" (thus less important) groups compared to "large" (thus more important) ones.

Table 1. Comparison of the four clustering approaches

Indicator	Class (CBC)	Owner (CBO)	Interactions (CBI)	Interactions with-Constraint on Actors (CBI-CA)
Total intra-cluster value (INTRA)	189	155	250	227
Averagenumber of actors	3.4	1	3.25	2.9
StdDev on number of actors	2.5	0	1.48	0.78
Maximum number of actors	7	1	6	4
Averagenumber of risks for eachactor	2.3	5.4	1.65	2.05
Mean cluster value	37.8	14.1	31.3	28.4
Stddev on cluster value	48	27	33	35
Mean cluster size	9.2	5.1	5.5	5.5

Stddev on cluster size	6.5	7.2	3.3	3.9
Number of riskswithin clusters	56	56	44	44

We still have to test other configurations, and especially to make a balance between the amount of interactions between risks and the number of assignments and size of groups. The possible correlation between these two last parameters will be analyzed in further work, since reducing the number of different risk owners in each cluster may be under certain conditions equivalent to reducing the number of cluster assignments for each actor.

Frequency analysis

In order to analyze the robustness of the proposed organization, different calculations have been run with $ClusterSize_{max}$ varying between 6 and 10, and with different configurations for a given vector ClusterSize. For instance, for $ClusterSize_{max}$=10, it is possible to test a five cluster configuration with each size of 10, or to test an eight cluster configuration with two clusters of 10, two clusters of 9, and so on. For each configuration, the calculation time has been recorded. Then, the frequency indicators are calculated and put in the frequency matrix, shown in Figure 5, which gives information about the robustness of this decision (the cells are colored to reflect the frequency values). A discussion is introduced with the decision-maker considering the proposed configuration and the complementary robustness analysis given by the frequency matrix.

Figure 5. Frequency Matrix built with the different tested configurations

The first conclusion is that the highest values are obtained for the biggest $ClusterSize_{max}$. This is essentially due to the presence of positive values only, and to the presence of enough non-null values in the original matrix (no saturation). Second, for a given $ClusterSize_{max}$, the best configuration is the one where the most clusters are fulfilled (their size being equal to $ClusterSize_{max}$).

But, it has to be noticed that in some cases, we found clusters with two or more independent sub-clusters. This means that in terms of clustering value, it does not bring anything, although in terms of human group coordination, it brings together people who do not have interactions. It can then be counterproductive to "artificially" group people with not enough reasons to do it. This is why it is not recommended to consider the merging of smaller clusters to make a team.

Except for some risks, the frequency of the chosen clusters is good enough to validate this solution. Some risks inside a cluster do not have a strong frequency index. Some risks outside a cluster have a strong frequency index with that cluster. But, the majority of proposals are validated by the frequency index. This means that it seems to be useful for future works as a pre-assignment technique in order to run more sophisticated optimization algorithms and software on a reduced problem. The clusters are partially sensitive to initial configuration parameters, but the majority of the solution is stable. This permits to be more confident with the solution.

Implications for managers

If management has a strategy to achieve early integration of risk owners and risk response decisions in order to detect and to mitigate potential propagation phenomena, then the use of this approach has to be done from the very beginning of the project. As a project is dynamic, whether in its objectives, components or context, this approach has to be used very early in the process, but also at different occasions and situations during the project. To enable appropriation of the approach, managers have to be committed to the both technical aspects, matrix-based modeling of risk network complexity and optimization-based decision-making. They have to be convinced and to create a context where the technical methodologies associated with the approach are understood, accepted and approved by engineers and managers.

In addition, the output of the approach indicates how the risk management structure needs to be changed, more precisely to be completed with a complementary and temporary task force-based organization, in order to create prerequisites for better communication, coordination and integration between project risk owners. In our case, managers including project managers and project office members have been at the origin of the work, not the operational risk owners.

Then, the support from top management was present, but the actors involved operationally in the process had to be convinced, with two main issues, the interest and the difficulty / additional energy. First, we assisted the process of capturing data and running calculations, explaining the concepts and involving the actors, but remaining leaders of the process. Second, the outputs of the first proposed configurations showed potential phenomena that corresponded to the experience of some risk owners, who declared that our highlighted risks seemed to be closer from reality (or at least what they lived before). This means that they trusted our proposal and found a potential interest to applying it.

At the end of the process, the approach received support from risk owners, project office members (in charge of proposing and deploying methods for projects) and top managers. Of course, some improvements were asked, whether to get the possibility to be more precise on the definition of the desired configuration (to put more parameters in the model), or to simplify some aspects of the approach (particularly for explanation or training, and more generally for appropriation by company members without the participation of researchers).

Conclusions

This paper presents an innovative risk clustering approach for efficient project risk management. The methodology enables comparisons between several possibilities for grouping risks in a project using several indicators: the total value of interactions inside the clusters and the structure of the clustering solution, in terms of cluster size, cluster value and cluster human composition. Our aim is to provide the decision-maker with complementary classifications which with the existing ones give powerful insights on the reality of complex phenomena in the project.

Since the clustering approach encourages people to meet together and communicate/ coordinate better, we consider that the overall communication / coordination performance is proportional to the performance of our algorithm. Indeed, the amount of interactions within the clusters (which is maximal) is a factual parameter. It determines a maximum potential for communication and coordination within clusters and a minimum risk of non-communication and/or lack of coordination at the interfaces between clusters.

However, even though the clustering decision can be more robust using the frequency approach we propose, this potential should be confirmed during the meetings and the day-to-day management of the project. If people are unable to agree and to coordinate, this will remain an untapped potential. It therefore refers to other aspects, such as the possible assignment of relevant Risk Cluster Owners, the use of meeting conducting techniques, collaborative decision-making techniques, general team management, etc.

In particular, the composition of the group (number of different actors, differences in terms of skills, background, hierarchical position, and experience) has to be carefully analyzed in order to increase the success probability of this heterogeneous but interrelated cluster. This is what we address in this paper, and further works will try to tackle more globally the assignment of actors to clusters, in terms of individual and collective parameters.

In the end, it is difficult to propose an objective measure of what we call the organizational capacity to cope with complexity, notably because it is a potential capacity. However, what is particularly important is that the risk of bad communication at interfaces is effectively reduced, since its probability decreases. There are less possible non communication situations and the ones that are remaining are the less important ones (regarding their occurrence probability).

The case study which is presented in the paper corresponds to a large project, which mainly includes aspects of civil work and design engineering. We think that the application field has an influence on the nature and number of interactions between risks. When testing the approach on several cases, we saw some differences between construction projects, new product development projects and musical show production projects. Even if the structure of project risk lists may vary (size of the list and density of the interactions between the risks), the clustering method does not depend on the application field and general conclusions about it can be extended to any domain. This could be an opportunity for future works to implement risk clustering on existing complex systems. Last, it could be worthy to assist the decision-maker to specify the desired configuration and to analyze the sensitivity of the clustering solution to this initial configuration.

References

Allan, N. and Yin, Y. (2011), Development of a Methodology for Understanding the Potency of Risk Connectivity. Journal of Management in Engineering, 27(2), 75–79.

Baccarini, D. (1996),The concept of project complexity – a review. International Journal of Project Management, 14(4): p. 201-204.

Camastra F. and Verri A. (2005), A novel kernel method for clustering, IEEE Transactions on pattern analysis and machine intelligence, Vol 27:5, 801-804.

Chen, S., Fath, B.D. and Chen, B. (2011), Information-based Network Environ Analysis: A system perspective for ecological risk assessment. Ecological Indicators, Volume 11, Issue 6, pp.1664–1672

Chen, S.-J., and Lin, L. (2003), Decomposition of interdependent task group for concurrent engineering.Computers and industrial engineering, Volume 44, Issue 3, pp. 435-459.

Chen, Y., Zhang, Y., Liu, J., and Mo, P. (2012),Interrelationships among Critical Success Factors of Construction Projects Based on the Structural Equation Model. Journal of Management in Engineering, 28(3), pp. 243–251.

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.

Eckert, C., J. Clarkson and W. Zanker. (2004), Change and customisation in complex engineering domains,Research in Engineering Design 15:1-21.

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): p. 1-13.

Fang, C., Marle, F., Zio, E. and Bocquet, J.C. (2012), Network theory-based analysis of risk interactions in large engineering projects. Reliability Engineering and System Safety, 106, 1-10.

Filippone M., Camastra F., Masulli F. and Rovetta S. (2008), A survey of kernel and spectral methods for clustering. Pattern recognition, 41.

Freeman, L. (1977), “Set of measures of centrality based on betweenness”. Sociometry 40:35-41.

Girvan M. and Newman M. (2002), “Community structure in social and biological networks”. Proceedings of the National Academy of Sciences (PNAS). doi: 10.1073/pnas.122653799, June 11, 2002 vol. 99 no. 12 7821-7826

Gomez C., Sanchez-Silva M. and Duenas-Osorio L. (2011), Clustering methods for risk assessment of infrastructure network systems. Applications of Statistics and Probability in Civil Engineering. 1389-1397.

Haponava, T. and Al-Jibouri, S. (2012), Proposed System for Measuring Project Performance Using Process-Based Key Performance Indicators. Journal of Management in Engineering, 28(2), 140–149.

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

Kloss-Grote, B. and M. Moss (2008), How to measure the effectiveness of risk management in engineering design projects ? Presentation of RMPASS: a new method for assessing risk management performance and the impact of knowledge management – including a few results. Research in Engineering Design 19:71-100.

Kwak Y. and F. Anbari (2009), Analyzing project management research: Perspectives from top management journals. International Journal of Project Management. Volume 27, Issue 5, July 2009, Pages 435–446.

Laurikkala H., Puustinen E., Pajarre E. and Tanskanen K. (2001), Reducing complexity of modeling in large delivery projects. International Conferenceon Engineering Design. Glasgow, Scotland.

McQueen J. (1967), Some methods for classification and analysis of multivariate observations. Computer and chemistry, Vol 4, 257-272.

Marle, F. and Vidal, L.A. (2008), Potential Applications of DSM Principles in Project Risk Management, Proceedings of the 10th International DSM Conference, 157-168

Marle, F., L.A. Vidal et J.C. Bocquet (2013), Interactions-basedriskclusteringmethodologies and algorithms for complexproject management. International Journal of Production Economics, Vol.142 Issue 2 pp 225-234.

Marle, F. and L.A. Vidal (2011), Project risk management processes: improving coordination using a clustering approach. Research in Engineering Design, Vol. 22, Issue 3, pp. 189-206.

Ng A., Jordan M. and Weiss Y. (2001), On spectral clustering, analysis and an algorithm. In Advances in Neural Information processing systems, 14.

PMI, S.C. (2008),A guide to the project management body of knowledge (PMBOK), 2008. (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): p. 53-66.

Schaeffer, S.E. (2007), Graph clustering. Computer Science Review I, 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, p.888-905

Simon, H. (1981), The Sciences of the artificial. Cambridge The MIT Press.

Tatum, C. (1989), Managing for Increased Design and Construction Innovation. Journal of Management in Engineering, 5(4), 385–399.

Van Dongen S. (2000), Graph clustering by flow simulation. Ph.D. thesis, University of Utrecht, Netherlands.

Vidal, L.A., Marle, F. and Bocquet, J.C. (2009), Interactions-based clustering to assist project risk management. Proceedings of the 17th International Conference on Engineering Design (ICED) Stanford USA.

Vidal, L.A., F. Marleet J.C. Bocquet (2010), Using a Delphi process and the Analytic Hierarchy Process (AHP) to evaluate the complexity of projects. Expert Systems With Applications, Vol. 38, Issue 5, pp. 5388-5405.

Wasserman S. and Faust K. (1994), Social Networks analysis: methods and applications. Cambridge University Press.

Xue, X., Shen, Q., and Ren, Z. (2010), Critical Review of Collaborative Working in Construction Projects: Business Environment and Human Behaviors. Journal of Management in Engineering, 26(4), 196–208.