Alma Mater Studiorum - University of Bologna
Department of Statistics 'Paolo Fortunati'

Ph.D. Program in Statistical Sciences
Curriculum in Statistics and Agricultural Economics
XXVIII Cycle

# Detection of communities of agents interacting through regional innovation policies

Ph.D. Candidate: Riccardo Righi
University of Bologna, Italy

Tutor: Professor Cristina Brasili
University of Bologna, Italy

Co-Tutor: Professor Margherita Russo
University of Modena and Reggio Emilia, Italy

Coordinator of the Ph.D. Program: Professor Alessandra Luati
University of Bologna, Italy

Settore Concorsuale 13/A2
Settore Scientifico Disciplinare Secs - P/02
Esame Finale Anno 2016

To my dear parents

No day goes by without my thinking
how fortunate I have been

## *Abstract*

The detection of communities of agents that interacted over time through regional innovation policies is analyzed through the application of three methodologies: Clique Percolation Method (CPM) by Palla *et al.* (2005), Infomap by Rosvall and Bergstrom (2008), and Dynamic Cluster Index analysis (DCI) by Villani *et al.* (2013). The case study regards the policy interventions implemented by region Tuscany (Italy) in 2000-2006 with the aim of supporting innovative network projects among local actors. In a context of analysis centered on such a complex object as innovation, and affected by discontinuous temporal dynamics and changing configurations of partnerships of agents, the three methodologies are applied to investigate different specific aspects of community organizations aimed at developing innovative activities. For every methodology three models are developed. In CPM, the elaboration of three models following the observation of the features of all possible partitions makes it possible to overcome the problematic definition of the value of $k$. In Infomap, the observation of the chronological order in which funded projects were carried out is used to impose different restrictions on the circulation of simulated flows. Finally, the application of DCI analysis to a socio-economic context is developed through the elaboration of different variables describing agents' behavioral profiles, and through an original contribution in using a cluster analysis aimed at coping with the large quantity of results that the algorithm produces. The investigation of relational structures (through CPM), of shared processes (through Infomap) and of integrated behaviors (through DCI analysis) allowed the identification of communities that reveal, respectively, meaningful characterizations in terms of agents' participations in specific waves of the policy, of agents' participations in projects operating in particular technological domains, and in terms of agents' institutional typologies.

# *Abstract*

Per individuare comunità di agenti che nel tempo hanno interagito in politiche regionali a sostegno dell'innovazione, si propone l'utilizzo di tre metodologie: Clique Percolation Method (CPM) (Palla *et al.* 2005), Infomap (Rosvall e Bergstrom 2008), e la Dynamic Cluster Index analysis (DCI) (Villani *et al.* 2013). Il caso di studio riguarda la serie di politiche messe in atto dalla regione Toscana, nel ciclo di programmazione 2000-2006, con lo scopo di sostenere progetti di reti innovative nel territorio.

Nell'analisi di un contesto che riguarda attività innovative, caratterizzato da forti discontinuità temporali nell'implementazione delle politiche e da mutevoli configurazioni nelle collaborazioni, sono state applicate le metodologie citate al fine di indagare tre specifici aspetti che caratterizzano le comunità di agenti con riferimento alla capacità di sviluppare processi innovativi.

Lo studio delle strutture relazionali presenti (attraverso il CPM), dei processi di interazione osservati (attraverso Infomap), e dell'integrazione dei comportamenti degli agenti (attraverso la DCI analysis) hanno condotto all'elaborazione di tre modelli di analisi distinti per ciascuna di queste metodologie.

Nell'ambito del CPM, la problematica definizione del valore di *k* è stata affrontata attraverso l'approfondimento delle caratteristiche delle possibili partizioni. Per l'applicazione di Infomap sono state elaborate simulazioni di flussi informativi in grado di tenere conto della sequenza temporale dei progetti finanziati. Infine, nell'ambito del DCI, un primo processo esplorativo, necessario per comprendere come applicare in modo coerente tale metodologia ad un contesto di tipo socio-economico, è stato seguito da due ulteriori modelli in cui l'originale introduzione di una analisi cluster ha consentito di gestire l'enorme mole di output prodotta dall'algoritmo. I risultati mostrano, rispettivamente, partizioni con comunità caratterizzate in termini di partecipazioni a specifici bandi (CPM), in termini di partecipazioni a progetti in specifici ambiti tecnologici (Infomap), e in termini di tipologia degli agenti coinvolti (DCI).

## *Acknowledments*

Firstly, I would like to express my sincere gratitude to Professor Margherita Russo whose expertise, understanding and patience have accompanied me unfailingly through the final part of my Ph.D. I am equally grateful to all the members of the research group I was given the pleasure and honor of joining. The passionate involvement in the research and their intellectual curiosity has been a constant source of inspiration to me.

My thanks go to the professors of the Department of Statistics at the University of Bologna, in particular my tutor Professor Cristina Brasili and the coordinator of the Ph.D. program Professor Alessandra Luati, for their assistance and understanding throughout. A big thank you to the staff members of the Ph.D Office, in particular Dr. Silvia Superbi, for the willingness and the kindness they showed, and a big thank you also to my fellow students on the course whose presence and friendship I have constantly appreciated.

I would like to thank all the professors and staff of the Department of Economics at the University of Modena, in particular Professor Paola Bertolini and Professor Enrico Giovannetti, without whose encouragement and motivation I would not have considered doing this doctorate, and Professor Paolo Silvestri, Dr. Antonella Cavallo, Dr. Sara Colombini, Dr. Francesco Pagliacci and Dr. Simone Righi, for the kindness they have shown and the support they have given through all stages of the project, and finally a special thanks to Professor Jane Stevenson for her invaluable assistance in the realization of this thesis.

To conclude, my profound thanks go to all of my family, to my friends, especially Marcello, and to all those who have stood by me through thick and thin. These three years have been intense, and it is thanks to their support and encouragement that I have been able to keep on track. It has been a thoroughly enriching experience.

# Contents

# Introduction

In mainstream economics literature, innovation has always had a particular place: a determining variable which cannot be controlled and whose appearance on the path of economic growth is relegated to being an 'event' (Solow 1956, 1957; Swan 1956); or an endogenous variable strictly modeled by one (or some) parameter(s) (Romer 1986; Lucas 1988; Rebelo 1991). The development of theories capable of considering the time dimension in a more articulated way, and capable of introducing the determinant concept of 'processes' in the study of economic activities (Hicks 1970; Georgescu-Roegen 1971) sparked the critical reviewing of mainstream growth theories and gave space to new considerations concerning the conceptualization of innovation processes (Amendola and Gaffard 1988, 1998). The academic discourse of economics reached the awareness that innovation was far from being an exogenous appearance or the product of some linear functions including some technological parameter. Thus, with the goal of deepening the comprehension of a phenomenon whose architecture was being seen more and more as articulated and complex (although complexity is not a synonym of impossibility nor of randomness), at the beginning of the 90's new contributions (Freeman 1988; Lundvall 1988, 1992; Nelson 1988, 1993) started to investigate the role of the circulation of information and of interactions among agents (people, enterprises and institutions) in the context of the flourishing of innovative processes. While the main aim of these authors was to study the ability of national economic systems to innovate, the introduction of such an approach in the literature fostered the investigation of the interactive dynamics that characterize the rise and the development of innovation processes at a micro-level (Lundvall 1985; Freeman 1991; Mowery and Teece 1996; Russo 2000) and at a meso-level of analysis (Saxenian 1994; Breschi and Malerba 1997; Malerba and Orsenigo 1997, 2002). As the interest in network dynamics and in the development of innovation increased, also institutional policy makers began to consider the relation between these two elements (European Council 2000; Audretsch 2002; European Commission 2003).

An inspirational contribution for the conceptualization and the investigation of innovation was given by David Lane (Lane and Maxfield 1997, 2005; Lane 2011) who outlined a theory capable of describing the most important entities and the most important phases that characterize the process of innovation formation. Far from looking for predictability, Lane's theory takes roots in the historical observation of the flourishing of innovation and, from that, highlights the fundamental patterns that typically occurred. Focusing on how the processes that led to innovations were intertwined, he was able to make some acute elements emerge. Even if for an ontological reason in Lane's approach the final objects (the results of innovation processes) remain unpredictable, thanks to his contribution fascinating paths of investigation have been opened up.

In fact, with Lane's theory as a background and moving towards the investigation of how innovation processes spread, important studies have been developed by a group of researchers of the Department of Economics, the Department of Communication and the Department of Physics, Informatics and Mathematics of the University of Modena and Reggio Emilia. The partnerships of these collaborations have produced several contributions

(Russo 2000, 2005; Lane 2002, 2004; Lane *et al.* 2005; Rossi 2007; Lane and Maxfield 2009; Serra *et al.* 2009, Villani *et al.* 2009; Russo and Whitford 2009; Russo and Rossi 2009a; Sardo 2009; Rossi *et al.* 2010; Bonifati and Villani 2013). Even if ontological or phenomenological or epistemological aspects have been investigated depending on personal areas of expertise, an important line of research which pointed to the comprehension of innovative processes has emerged and continues to evolve. Among these partnerships, a research team has focused on the economic investigation of the local and regional development of innovative processes, with particular regard to their relation with policy interventions. This research group, which has its fundamental references in Margherita Russo[1], Annalisa Caloffi[2] and Federica Rossi[3], is the one I had the opportunity and the privilege to join for the work of this thesis.

One of the most recent and prolific focuses that this group has developed concerns a peculiar and incredibly interesting case study regarding an entire policy cycle, which took place in Tuscany, Italy, in the context of the programming period 2000-2006, in support of innovation. Even if several studies have been carried out on this specific case (Russo and Rossi 2005, 2007, 2009b; Caloffi and Rossi and Russo 2011, 2013, 2014, 2015), I gave my personal contribution with a further development which specifically concerns the investigation of communities among those agents that participated in the mentioned policies. In a case study regarding a situation in which (i) all the policies were devoted to the support of innovation processes and (ii) all the policies allowed the granting of funds exclusively to network projects, I explored ways to group agents on the basis of the behaviors they had and on the basis of the interactions they produced. Referring in particular to the conceptualization of 'organizations' developed by Lane (Lane 2011), where organizations are intended as groups of agents characterized by the presence of relational *structures* and shared *processes* and common *functions*, I developed an analysis that took three directions, each of which regarded the application of a specific methodology, one for each of these aspects. The three methodologies considered to investigate the presence of groups of agents were: (i) Clique Percolation Method by Palla *et al.* (2005), (ii) Infomap algorithm by Rosvall and Bergstrom (2008), and (iii) Dynamic Cluster Index analysis by Villani *et al.* (2013, 2015). The observation of a situation in which agents always have to constitute partnerships aimed at the development of innovation projects (the considered policies allowed exclusively the granting of funding to partnerships of agents), has to be regarded as an incredible opportunity to investigate some specific aspects of a phenomenon whose flourishing is more and more thought of, by the specific economic literature, as being deeply rooted in the interactive dynamics that occur among agents. Thus, what seemed to me to be important was to investigate if agents (which in this context are enterprises, Universities, Chambers of Commerce, Trade Associations, Service Centers, Business Services Enterprises and other typologies of economic institutions) interacted among themselves, through their participations in projects in a series of public policies that overall took seven years, to give

[1] Professor at the Department of Economics, University of Modena and Reggio Emilia

[2] Lecturer at the Department of Economics and Business, University of Padova

[3] Lecturer at the Department of Management, School of Business, Economics and Informatics, Birkbeck, University of London

rise to entities which can go beyond the boundaries of the single projects and in which competences and connections created architectures with potentiality to foster the flourishing of innovation processes. Thus, the main and most general research question to which I have tried to give an answer is

*how did agents interact among themselves during this cycle of policies?*

Starting from this point, the consideration of the specific features of the case study under analysis led to new and more detailed research questions and I developed an approach which mainly moved towards the deepening of the three methodologies mentioned above. Hereunder, the structure of the thesis is presented.

In the first chapter, I explain how the analysis was approached. After the consideration of the specific and determining features of the case study, the theoretical framework upon which the analysis is based is described. Then, I discuss how some highlighted elements determined the formulation of research questions and, to conclude, I explain how methodologies were selected in accordance with the previous considerations.
In the second chapter, I present the structure of the available dataset.
In the third chapter, I deal with the problematic application of Clique Percolation Method by Palla *et al.* (2005), the first methodology that I used for the analysis. In particular, the delicate definition of the value of *k* is discussed and the results produced by CFinder software are presented. Then, I describe how partitions with similar features can be grouped with respect to three ranges of *k* and, after the comprehension of the main and typical features of each of these groups, I select three specific values of *k* which define three different models for the analysis: CPM_k05, CPM_k12 and CPM_k18.
In the fourth chapter, I introduce the main concepts of the two-level Infomap algorithm by Rosvall and Bergstrom (2008, 2014), the second methodology that I took into account, and in particular I describe how this procedure can be used in order to apply second order Markov models. Then I explain how it was possible to configure available data to apply different temporal conditions, and hypotheses about the reconstruction of possible circulations of information among agents (through participations in common projects) are discussed. Finally, I describe how different hypotheses concerning the exchange of information among agents led to the configuration of one memoryless model and two models of second Markov order: MKV1, MKV2_AS and MKV2_DS.
In the fifth chapter, I start by introducing the studies carried out by Tononi *et al.* (1994, 1996, 1997, 1998) about the detection of functional groups of neurons inside the brain area. After the explanation of some concepts taken from the discipline of information theory, I introduce the Dynamic Cluster Index analysis by Villani *et al.* (2013, 2015), the third methodology that I dealt with. Here, one of the most important points of the thesis is described: the attempt to apply this new methodology in the economic research field. All theoretical and practical difficulties are discussed and, after that, the solutions I adopted are illustrated. In particular, a consistent part of the chapter is dedicated to the description of the first process of analysis: BOOL_1. Through this exploration, many problematic aspects emerged giving me the possibility to manage more coherently the two following processes of

analysis: BOOL_2 and ∂LoA. From the point of view of the settings, these two models are equal, since the procedure developed, thanks to the attempts made in the context of process of analysis BOOL_1, was applied to both in the same manner. However, these two models differ from each other because of different informational bases. The first, BOOL_2, involves the use of boolean variables, as were used in the explorative process of analysis BOOL_1, while the second, ∂LoA, involves the use of variables regarding the variation over time of the number of projects in which agents were participating.
In the final chapter, results are presented. I discuss the most evident features of the partitions produced by each model and considerations about these communities are made with particular attention to the characteristics of agents which belong to them. To this end, the observation of the statistical significances between the expected frequency and the observed frequency (adjusted residual) in meaningful contingency tables are used as the principal criterion to highlight specific characterizations of these communities. Simultaneously, I consider the community structure that every partition depicted, investigating in particular the presence of overlaps among communities. I make observations about the comparison among the models that refer to the same methodology, and I try to investigate to what extent they produce similar partitions (or not). I also provide considerations about the presence of a continuity in the applications of the three models which have been developed within the context of each methodology. Finally, I try to move the final observations beyond the boundary of the single methodologies and I recall the theoretical considerations taken into account in the first chapter. Following Lane's interpretation of 'organizations' - which are interpreted as groups (of agents) characterized (i) by a structure, (ii) by the ability to manage processes, and (iii) by the presence of a common function (which is counterposed to random activity) - I investigate the presence of groups (of agents) which simultaneously respond to these three characteristics. I compute the Jaccard similarity coefficient of all possible combinations of three communities as follows:

- the first community is one of the communities detected with CPM;
- the second community is one of the communities detected with Infomap;
- the third community is one of the communities detected with DCI.

In this way, I make further considerations about the overlaps among communities which, since they were detected with methodologies that moved from different perspectives, should put in evidence specific topics of investigation.

# 1. The development of the approach

The case study that is taken into account in this work presents specific features that have to be described immediately since the configuration of the entire approach to the analysis depended on them. While the available informational basis is described in the next chapter, here the features that characterize the cycle of policies that Region Tuscany promoted from 2000 to 2006 in order to support innovation processes are described and, along with the introduction of these elements and along with the development of crucial considerations, the approach to the analysis is explained.

## 1.1. The case study

First of all, the most important point concerns the presence of two characteristics which represent the most peculiar elements of this case study:

- the entire cycle of policies granted funds to projects whose objective was the development of innovative processes;
- the entire cycle of policies granted funds to network projects[4] .

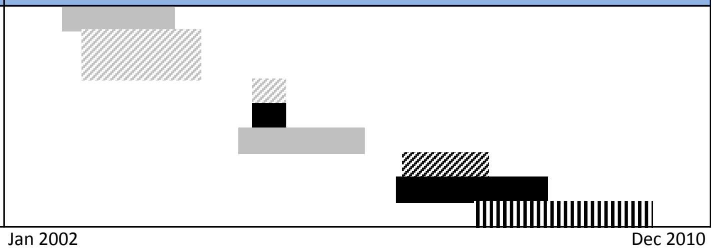
These two features made it possible to specifically orient the analysis to a definite object: the detection of group of agents that, interacting among themselves, tried to give rise to innovation processes. The interesting alignment between the framework which, according to a specific theory of innovation, occurs in the processes of the formation of innovation and the two main features of these policies is crucial in the development of the whole approach. This aspect, which is discussed more thoroughly at the end of the paragraph 1.2 and in paragraph 1.3, must always be borne in mind. In the same way, it is important to describe how these policies were characterized, since the emergent features are crucial in the phase of selection of the methodologies and in the phase of the development of the models.
The considered cycle of public policies was promoted by the Region Tuscany and was composed, at the end of it[5], of nine different waves[6]. The period over which these waves were promoted started in 2000 and went on until 2006. In reality, the period 2000-2006 is only the period to which the policies refer, and the effective implementation of projects

[4] The considered policies allowed exclusively the granting of funds to projects whose promoter was a group of agents (a partnership).

[5] When the policy cycle started, institutions did not have programmed the number of interventions. These decisions were taken during the development of the policies.

[6] Every policy referred to a specific programme. Thus the combination of the kind of programme and the year in which it was developed determined the nine different waves. In the list that is present in this footnote, for every year the programs supported are presented and (in brackets) is indicated the name that has been used to refer to the specific wave.
In 2002 three different waves were promoted: PRAI-ITT (2002_ITT), 171 (2002-171) and 172 (2002_172). In 2004: 171 (2004_171) and 171A (2004_171A). In 2005: 171 (2005_171). In 2006: PRAI-VINCI (2006_VIN). In 2007: 171 (2007_171). In 2008: 171 (2008_171).

| Policy Wave | Date of beginning of the first funded project | Date of end of the last funded project | Gantt diagram by month |
|---|---|---|---|
| 2002_ITT | 30 Sep. 2002 | 29 Feb. 2004 | |
| 2002_171 | 31 Dec. 2002 | 30 Jun. 2004 | |
| 2002_172 | 31 Dec. 2002 | 30 Jun. 2004 | |
| 2004_171 | 28 Feb. 2005 | 01 Jun. 2005 | |
| 2004_171A | 01 Mar. 2005 | 01 Jul. 2005 | |
| 2005_171 | 31 Dec. 2004 | 31 Jul. 2006 | |
| 2006_VIN | 14 Feb. 2007 | 29 Feb. 2008 | |
| 2007_171 | 31 Dec. 2006 | 30 Nov. 2008 | |
| 2008_171 | 31 Dec. 2007 | 01 Mar. 2010 | |



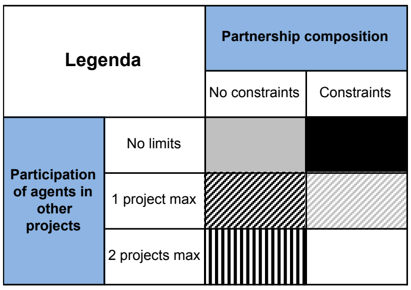




*Figure 1.1 (above). Gannt diagram (by month) representing the period in which the projects (supported by the different waves) were realized. For every waves was considered the date of beginning of the first (in chronological order) funded project and the ending date of the last project funded. These dates are shown beside the diagram. Below the diagram are indicated the month representing the first column of the diagram and the month representing the last one.*

*Figure 1.2 (left). Combination of constraints that were imposed to the participations to every wave. The textures used to identify the different combinations are used to characterize the waves in the Gannt diagram.*

occurred in the period 2002-2010[7]. Considering this time dimension, it is important to observe that waves were not uniformly distributed over time: they had different durations and they overlapped, producing periods in which no wave was active and periods in which three waves were simultaneously active. This fact, which can be easily observed in the Gannt Diagram (figure 1.1) immediately makes it possible to understand how the implementation of these policies was not constant over time. In addition to these elements, every wave presented specific features with regard to:

- the presence of constraints regarding the composition of the partnerships;
- the presence of constraints regarding the possibility of participating in more than one project in the context of the same wave;
- the technological domains in which projects were allowed to operate;
- the amount of financial resources made available;
- the percentage (on the basis of the costs) of the grants of funds to each single project.

The first two elements that are listed above have to be considered the most important elements with respect to the evaluation of admissibility of projects since, if not respected, the possibility of receiving funds was immediately excluded. The first one imposed that the partnerships have to be composed (i) either of a minimum number of agents of a specific typology, (ii) or of a minimum number of agents in which some specific typologies have only to be present. On the other hand, the second constraint prevented agents' participation

[7] The beginning of the first funded project is dated 30 September 2002, and the end of the last funded project is dated 1 March 2010. This fact does not cause any important variation with respect to the analysis, however, it is important to describe it.

| Technological Domains | Policy Waves 2002_ ITT | 2002_ 171 | 2002_ 172 | 2004_ 171 | 2004_ 171A | 2005_ 171 | 2006_ VIN | 2007_ 171 | 2008_ 171 | TOT. |
|---|---|---|---|---|---|---|---|---|---|---|
| Others | | | 1 | | | 1 | | 1 | 1 | **4** |
| Bio-Technologies | 2 | | 1 | | | 1 | | 1 | 1 | **6** |
| Organic Chemistry | 1 | | 1 | 3 | 7 | | 2 | 1 | | **15** |
| Geotermal Sciences and Biomasses | | | | 1 | 4 | | | | 1 | **6** |
| ICT - Multimedia | 6 | 1 | 5 | | | 21 | 7 | 24 | 15 | **79** |
| Mechanics Engineering | | | | 1 | | 4 | 2 | 3 | 2 | **12** |
| Multi Disciplinary | | | | | | 1 | | | 2 | **3** |
| Nanotechnologies | 1 | | | | 1 | 1 | | | | **3** |
| New Materials | | | | | 1 | 4 | | 1 | 1 | **7** |
| Optoelectronic | 4 | | | | 1 | 3 | 1 | 2 | 3 | **14** |
| NA | | | | 1 | | | | 8 | 10 | **19** |
| **TOTAL** | **14** | **1** | **8** | **6** | **14** | **36** | **12** | **41** | **36** | **168** |

*Table 1.3. Number of projects that were funded in the context of each wave, by the technological domain to which their objective refers to.*

| Wave | Number of funded projects | Number of non funded projects | Number of proposals | Average amount of resources available, per project and wave (in Euro) | Percentage of the amount of resources available per wave, up to the total amount of resources granted | Total amount of resources made available, per wave (in Euro) | Number of agents that participated in funded projects |
|---|---|---|---|---|---|---|---|
| **2002_ITT** | 14 | 22 | 36 | 329.821 | *12,55%* | 4.617.490 | 252 |
| **2002_171** | 1 | 1 | 2 | 1.465.820 | *3,98%* | 1.465.820 | 35 |
| **2002_172** | 8 | 8 | 16 | 193.810 | *4,21%* | 1.550.479 | 76 |
| **2004_171** | 6 | 22 | 28 | 62.910 | *1,03%* | 377.459 | 42 |
| **2004_171A** | 14 | 0 | 14 | 53.679 | *2,04%* | 751.500 | 70 |
| **2005_171** | 36 | 19 | 55 | 112.438 | *11,00%* | 4.047.785 | 833 |
| **2006_VIN** | 12 | 15 | 27 | 238.303 | *7,77%* | 2.859.640 | 80 |
| **2007_171** | 41 | 28 | 69 | 244.163 | *27,21%* | 10.010.692 | 333 |
| **2008_171** | 36 | 15 | 51 | 308.658 | *30,20%* | 11.111.678 | 282 |
| **TOTAL** | **168** | **130** | **298** | **219.003** | **100,00%** | **36.792.543** | |

*Table 1.4 (above). Descriptive statistics that regards the waves promoted by Region Tuscany during the programming period 2000-2006.*

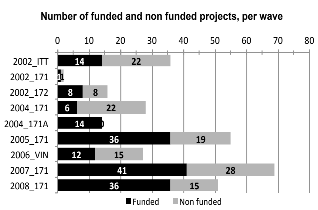



*Chart 1.5. Histogram showing the number of funded and non funded projects, per wave.*

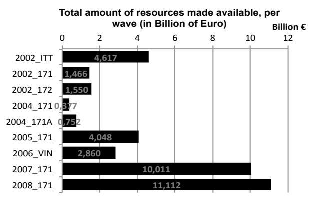



*Chart 1.6. Histogram showing the total amount of resources made available, per wave.*

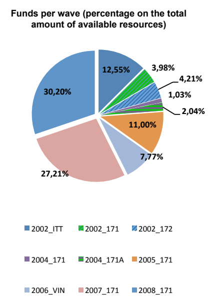



*Figure 1.7. Pie chart representing the percentages (up to the total that were made available) of the amounts of funds, per wave.*

in more than a pre-established number of projects. These two constraints were not applied in every wave and, when they were present, their combination was articulated in different ways (figure 1.2). In some cases these two constraints were both active, in other cases only one of them was applied, and in others there were none. This fact determines the first important element of heterogeneity among waves. While the lack of a constant presence over time has already been observed, here the large diversities in the constraints of each policy intervention have been highlighted. The third element which is listed above regards the presence of specific technological domains. Every wave, in fact, allowed the granting of funds in specific technological domains which, moving from wave to wave, were not always the same. What occurred is presented in the table 1.3. It can be observed how the distribution of the technological domains of projects is different in each single policy intervention. After the application of various type of constraints, this constitutes the second element of difference among waves. Finally, not all the projects that were admissible were also allowed to receive funds and what determined the final number of projects, which in the context of every wave were funded, was: (i) the general availability of resources; (ii) the maximum percentage of the covering of the costs of the projects.

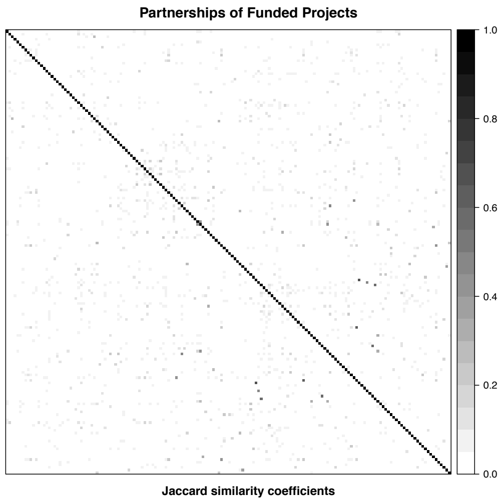
Thus, along with the other elements described above (constraints and technological domains), the availability of resources and the maximum coverage that could be guaranteed to projects determined a situation of intense diversity among waves. It can be observed (table 1.4, charts 1.5 and 1.6 and figure 1.7) that the number of funded projects, the number of non funded projects, the number of proposals and the amount of resources granted, vary greatly from wave to wave. Excluding the wave 2002_171 , which has only one funded project, what can be observed is:

- a range from a minimum of 6 projects funded (wave 2004_171) to a maximum of 41 projects funded;
- a range from a minimum average of 53.679 Euro per project (in wave 2004_171A) to a maximum average of 329.821 Euro per project (in wave 2002_ITT);
- a range from a total amount of 377.459 Euro of available resources over the entire 2004_171 wave to a total amount of 11.111.678 Euro of available resources over the entire 2008_171 wave;
- a range from a minimum of 42 agents involved (in wave 2004_171) to a maximum of 833 agents involved (in wave 2005_171).

Other examples could be given, but the situation has already been depicted sufficiently. Every wave presents peculiar features in terms of (i) the number of projects, (ii) the number of agents involved and (iii) the resources available.

To conclude the presentation of the case study, there is a final consideration that has to be made regarding the partnerships that were observed during the entire policy cycle. Agents that participated in more than one project (irrespective of the waves in which the projects were collocated) were not obliged to participate always with the same group of agents. This means that every observed partnership could be made up of a different combination of agents. Analyzing the Jaccard similarity coefficient between all possible couples of partnerships in terms of agents that are included in them (figure 1.8), it emerges that only

few cases show similar compositions and, to be more precise, only five couples of projects have at least 50% of agents in common (out of the union set of each couple, which is the total number of agents that participated in the two projects considered) and only two of them have a coefficient higher than 0,6 (respectively 0,66 and 0,63). Since 168 projects were realized, the possible couples of partnerships are 14.028 and the percentage of those that show similar structures in terms of agents is extremely low[8] (see figure 1.9). Moreover, the strong diversification in terms of partnerships can be observed also thanks to the representations of the graphs of all the waves (graphs 1.10). Attributing to every agent a constant position in all the graphs (and representing them only if in the corresponding wave they participate in at least one project), it can be immediately noticed to what extent each wave is different from the others in terms of partnerships' configurations.
The case study has been presented and its most important features have been highlighted. How the database was managed and which kind of information it contains is described in chapter 2. Here it was important to give a general description of the case under analysis, in order to comprehend methodological choices that were made with respect to theoretical aspects. In the next paragraph, I introduce some theoretical elements concerning the concept of 'innovation'. After describing the situation that was taken into account, a theoretical framework is needed to orient the development of the analysis.



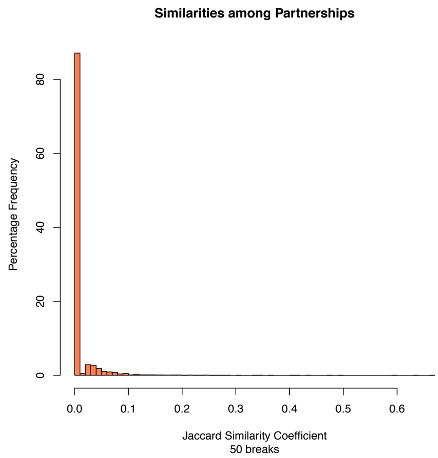
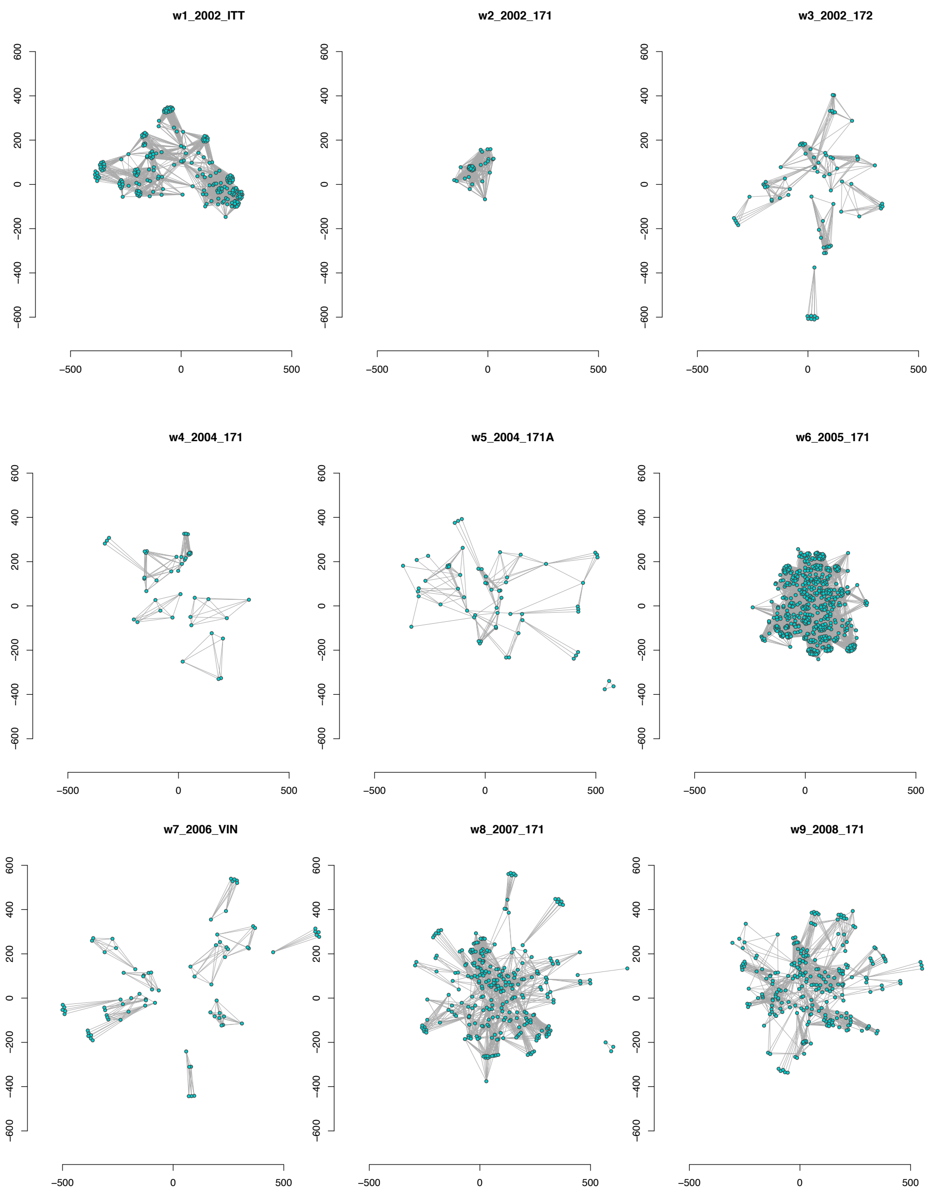




*Figure 1.8 (left). Matrix of Jaccard similarity coefficients calculated for all pairs of partnerships that were composed for the development of the projects (funded). The darker is the color, the higher is the ratio. Rows and columns represent the partenerships.*

*Figure 1.9 (right). Percentage frequencies of the Jaccard similarity coefficients calculated for all pairs of partnerships that were composed for the development of the funded projects (see figure 1.8).*

[8] Five couples of partnerships (those with the intersection that is at least the 50% of the union set) over a total of 14.028 possible pairs of them represents the 0,0356%.




*Graphs 1.10. Graphs representing the agents that took part in each of the nine waves. To every node is attributed a constant position in all the graphs, and they are represented only if in the corresponding wave they participate in at least one project. The edges represent the common participations in at least one project (in the context of the corresponding wave). All the graphs are plotted on the same scale. Fruchterman Reingold layout.*

## 1.2. Insights from innovation theory

Innovation processes, the delicate object under analysis, need to be investigated and comprehended in order to acquire those notions necessary to provide a useful theoretical framework for a coherent development of the analysis. The aim of this paragraph is to underline some elements, taken from David Lane[9]'s theory of innovation, which with regard to this thesis have been crucial in defining an appropriate way to analyze the specific case study. After having briefly introduced two concepts - *exaptive bootstrapping* and *generative relationships* - which are fundamental in this theory of innovation[10], the focus is directed to particular entities that deserve a special attention: 'organizations'.

First of all, the general domain which David Lane's innovation theory assumes as context of reference has to be introduced: the agent-artifact space. This specific dimension concerns the interaction between the two most important entities in Lane's vision, agents and artifacts. While the first represent all those entities that take action, the second concern those means that are produced and used by the first in order to "*wrest more usable matter and energy from their environments, allow more of them to live in a given environment, and sometimes allow them to live longer and dedicate more of their time and energy to reproduction*" (Lane 2011). Thanks to their ability to act, agents can produce and use artifacts which help them in their existence, and this capacity has to be regarded as the specific ability to handle an environment which has to be interpreted as a framework that changes as functionalities attributed to artifacts change[11]. The schematization of how functionalities change over time is provided by the representation of the so-called 'exaptive bootstrapping'[12] sequence, a series of phases that typically occur in the context of the flourishing of innovation. As described by Lane (Lane 2011) the phases of the exaptive bootstrapping are:

---

[9] David Lane was Professor (1976-1992) and Chairman of the Department of Theoretical Statistics (1989-1992) at the University of Minnesota. In 1992, he became Professor of Statistics in the Department of Political Economy at the University of Modena, of which he was Chairman (1998-2001). From 2001, he is member of the Faculty of Communications and Economics of the University of Modena and Reggio Emilia, where he currently serves (since 2003) as professor of economics. In addition, Lane was a member of the external faculty at the Santa Fe Institute from 1989 until 2009; he continues to serve on the Institute's Science Board and Editorial Board. He is a member of the Editorial Board of the journals Complexity and Journal of Evolutionary Economics.

10 The theory proposed by David Lane does not have as objective the prediction of the results of innovation processes. This, could be conceded as an oxymoron. Quoting Lane (2011): "From one point of view, a "theory of innovation" is an oxymoron. If, as many scientists believe, a theory is supposed to lead to verifiable predictions of the phenomenon under study, then a theory of innovation should predict innovations – which would mean the process leading to innovations the theory was meant to explicate is just an historical dead-end that could be replaced as innovation-generator by the theory itself! Of course, this is silly: the theory could illuminate aspects of the process without "predicting" the new artifacts that were the process outcomes of primary economic and social interest". The aim of the theory that I took into account for this work regards the the understating of how innovation processes develop, under which conditions they flourish and who are the entities that are involved.

[11] One of the core element of Lane's theory is the conceptualization of the space of interaction between agents and the context in which they live, a dimension in which agents are able to attribute new functionalities to artifacts.

[12] The term 'bootstrapping' refers to the fact that the sequence, which is made up of five stages, has a circular shape, since its first stage begins where the last finishes. On the other hand, the term 'exaptive' alludes to the capability to foresee and to anticipate attributions and functionalities that do not exist in the present environment, an element that has to be regarded as fundamental.

- New artifact types are designed to achieve some particular attribution of functionality.
- Organizational transformations are constructed to proliferate the use of tokens of the new type.
- Novel patterns of human interaction emerge around these artifacts in use.
- New attributions of functionality are generated - by participants or observers - to describe what the participants in these interactions are obtaining or might obtain from them.
- New artifacts are conceived and designed to instantiate the new attributed functionality. (Lane 2011)

The most important passage of this sequence lies between the third and the fourth stages[13]. Here the exaptation occurs; here agents anticipate new functionalities that can change the relation between themselves and the environment in which they live.

In the theory proposed by Lane, the capacity to generate new attributions (functionalities) depends on the interactions that occur in the agent-artifact space. These interactions have to be distinguished into two different types, cognitive and communicative, and, while the former determine the possibility to trace new routes in the agent-artifact space, the latter are needed to make these routes concrete[14]. Both these relationships are necessary in the process of formation of innovation described above, but the element which has not to be taken for granted (and which constitutes one of the most important points in Lane's conceptualization) is that they both involve interactions among agents. In fact, if on one hand the communicative processes obviously deal with an interaction among agents, on the other hand it is not obvious that the processes of development of new attributions (cognitive processes) have to deal with interactions involving a plurality of agents. Thus, Lane attributes to relationships among agents a generative potentiality[15] that has to be seen as the core element in the process of the flourishing of innovation.

Since innovation processes deal with interactions among agents, the last theoretical step is to define a concept that regards the grouping of agents interacting among themselves. David

13 Quoting Lane (2011): "Exaptation is the taking on of new functionality by existing structure. It happens between the third and the fourth stage in this process, whereby new attributions of functionality arise from observing patterns of interaction among agents and already existing artifacts"

14 Quoting Lane (2011): "The most important cognitive process in innovation is the generation of new attributions. Similarly, the most important communication processes involve the aligning of attributions among agents, otherwise the processes of recruitment, differentiation and coordination that underlie the collective action necessary to transform new attributions into new artifacts into new patterns of interaction cannot take place."

15 Quoting Lane (2011): "These new attributions arise in the context of a particular kind of relationship among agents, which we call generative. While the kind of ontological uncertainty that typically shrouds innovation processes makes it impossible to predict in detail what sorts of new attributions a relationship may generate, it still may be possible for agents to assess the generative potential of a relationship. This potential depends on five characteristics of the agents in the relationship and their modes of interaction with one another, and agents may not only infer the degree of these characteristics through their interactions, but may also act in such a way to increase the relationship's generative potential."

Lane identifies this concept with the term 'organizations', which in this specific context have to be intended as groups of agents that (interacting among themselves) are able to give rise to those processes that can transform the environment. Lane is very clear in identifying which elements characterize the entities 'organizations': "*Organizations are particular kinds of interacting entities, which can be characterized by three fundamental aspects: structure, function, and process*" (Lane 2011). The main features of these 'interacting' entities emerge: the presence of a (i) structure, (ii) the presence of processes, and (iii) the presence of a shared function. Lane gives a detailed explanation of what these concepts mean and of what they imply. Quoting Lane (2011):

> Structure is recursive: it has parts, which are themselves organizations. To describe an organization's structure, one must identify its parts, the interaction modalities among its parts, and the modalities through which the organization interacts with other organizations. Hierarchy and networks are key concepts in describing organizational structure: the former, because of recursivity; the latter, because organizations engage in recurring patterns of interaction, which constitute networks – with organizations as nodes and recurring interactions as links.
> The processes associated with an organization describe the transformations (in the organization of its world) in which the organization may participate. Processes are supported by structure. To enact any of its processes, some of an organization's parts must engage in interaction events, each of which requires some particular interaction modality. (…)
> The functions of an organization provide directedness to its actions, through their role in determining which processes the organization enacts, when the organization is in a context in which more than one process could be enacted. (Lane 2011)

Thus, the *structure* concerns the shape of an organization, with particular regard to the interaction modalities that its framework makes possible. *Processes* regard the activities which the agents that belong to the specific organization embark on. Finally, the *functions* have to be intended as the purposes which guide agents' actions. This theoretical framework is all that is needed to have points of reference thanks to which the approach to the analysis can be developed.

The case study that was taken into account regards a series of public policies that concerned the creation of innovation projects. The peculiar object in the focus of these policies (innovation) immediately needed to find a specific theoretical ground to host considerations for which there is little room in other theoretical frameworks. The introduction of some specific entities - agents, artifacts and the interactions that occur among them - allowed the definition of the agent-artifact space, a theoretical dimension characterized by an

environment which can be continuously redefined and reshaped by agents through the creation of artifacts and through the attribution of new functionalities to them.
Once the stages that characterize innovation processes had been highlighted (exaptive bootstrapping), it emerged how it is not only a matter of agents that interact with artifacts, but that above all it is a question of interactions among agents (generative relationships). Thus, as Lane depicted, the process of evolution of the environment - and so the process of transformation of the living conditions of agents - is necessarily carried on by collective action. In this context, what becomes important is how agents interact among themselves. As a result, a third entity emerged which first of all has to be recognized as fundamental in the whole theoretical structure, and which in turn has to be recognized as crucial in this specific analysis. In fact, the case study that I investigate regards agents that had necessarily to aggregate themselves in groups, in order to participate in the considered public policies. In other words, these agents had to constitute what can be seen as forms of organizations. That is the reason why to direct the focus on the concept of 'organization' seemed to me the best option to investigate the presence of groups of interacting agents. And to do this, I decided to approach the analysis with the intent to apply three different methodologies that could allow me to investigate those aspects of organizations that have been described in this specific literature regarding innovation (and it must remember that innovation is in fact the specific object of the considered cycle of policies). These concepts, which are *structure* and *processes* and *functions*, were the most important points of reference in the development of the analysis and, in particular, in the selection of the methodologies I applied. Other considerations more related to methodological aspects are made in the next paragraphs. However, the direction of the analysis has already been traced.

## 1.3. The research questions

Starting from the consideration of the specific features of the case study under analysis, the two elements that have immediately to be remembered to formulate appropriate research questions regard these aspects: (i) all the considered policies aimed to sustain innovation, and (ii) all the policies involved exclusively network projects. Only after having introduced the theoretical perspective developed by David Lane, at the core of which there is the concept of innovation intended as a process of interactions, does it becomes evident that an adequate development of the analysis has to pass through the consideration of how involved agents interacted. Thus, the first research question assessed by the working group was:

*how did agents interact among themselves during the cycle of considered policies?*

In order to answer this generic question, it was necessary to give it a more specific formulation, and this was done through the understanding of what characterized mostly the object under analysis.

### 1.3.1. The detection of communities

The consideration of Lane's theory about innovation permitted me to focus the attention on the relation between innovation and interactions among agents. In this theoretical framework, another fundamental concept emerged, which seems to be appropriate and helpful in guiding the analysis: the concept of organizations. The definition of the entities 'organizations' and the understanding of their importance in the context of innovation processes, led me to go beyond the generic intention to investigate how agents interacted among themselves. In the light of theoretical elements introduced, now the research question can be developed in a more detailed way:

*how is it possible to investigate the presence of communities of agents that were trying to produce innovative processes, in the context of the considered public policies?*

As can be noticed, the new specification regards the object that is investigated. While before the intent was to study generically 'interactions', now what is under analysis are 'communities'[16], entities that can be clearly thought of as parallel to the concept of organization developed above[17]. Thus, in order to conclude and disregarding these differences of language forms, it has become explicitly clear that what has to be analyzed are not interacting agents considered separately, but the presence of entities containing a plurality of agents. What is investigated are the entities that made interactions possible.

### 1.3.2. The time dynamics

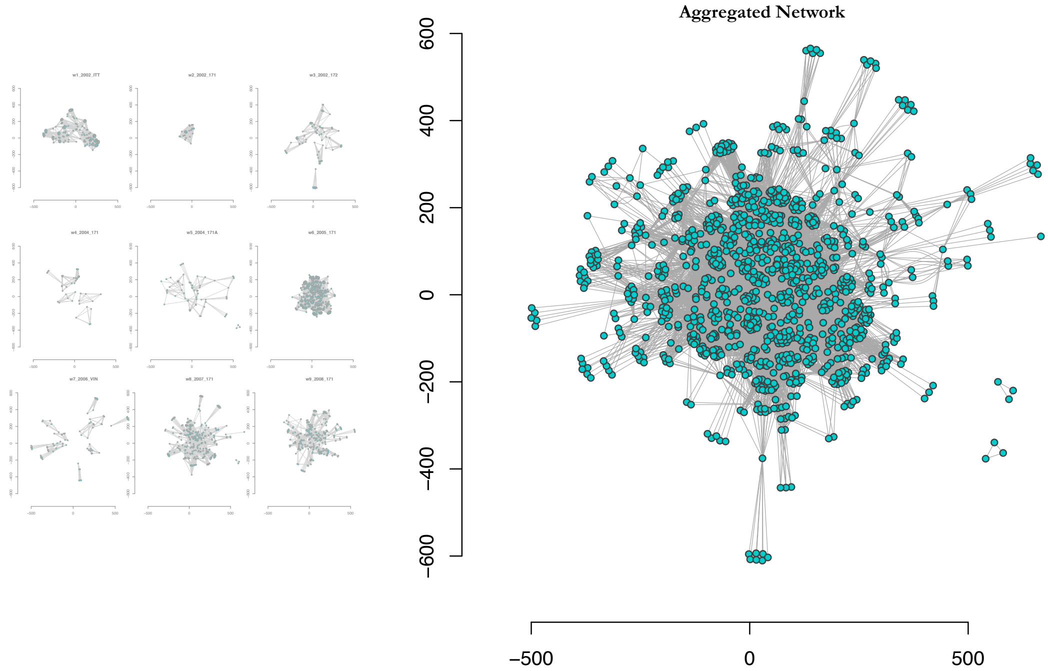
With regard to the structure of the policy waves, what was highlighted in paragraph 1.1 was that agents participated (i) discontinuously over the different waves, and (ii) with different configurations of partnerships. On one hand, the fact that the entire cycle of policies took 6 years immediately imposed the consideration in the analysis of time dynamics. On the other hand, the presence of a process of evolution of communities in the depicted sequence of considered policy interventions cannot be observed. Even if the nine policies can be ordered in time as separate instants, they do not show any kind of dynamics in terms of creation or evolution of possible partitions in which agents can be involved. This fact, led to the development of a new research question whose formulation is:

*how is it possible to investigate the presence of communities of agents that were trying to produce innovation processes, in a context in which the observed networks' configurations change dramatically over the nine different waves of these public policies?*

---

[16] Santo Fortunato (2010) uses in his definition of 'community' the word 'organization': "One of the most relevant features of graphs representing real systems is community structure, or clustering, i. e. the organization of vertices in clusters, with many edges joining vertices of the same cluster and comparatively few edges joining vertices of different clusters. Such clusters, or communities, can be considered as fairly independent compartments of a graph, playing a similar role like, e. g., the tissues or the organs in the human body". (Fortunato 2010)

[17] Since the entity 'organization' - defined in an elevated theoretical framework - regards the definition of groups of agents interacting among themselves, it seemed to me adequate to associate this concept to the one of 'communities'. It is only a matter of research fields and in fact, if on one hand in innovation theory David Lane speaks about organizations, on the other the social network analysis literature uses the concept of communities.




*Graphs 1.11. On the left are the graphs representing the agents that took part in each of the nine waves are plotted. To every node is attributed a constant position in all the graphs, and they are represented only if in the corresponding wave they participate in at least one project. The edges represent the common participations in at least one project (in the context of the corresponding wave). All the graphs are plotted on the same scale. The aggregation of the nine networks produced the graph plotted in the right. The 1.1121 agents (nodes) that participated in at least one project and all the participations in common projectes (edges) that occurred are simultaneously represented. Fruchterman Reingold layout.*

To answer this last question, I looked at the available networks from a different perspective. Since to look at the sequence of waves in order to investigate how these networks evolved and changed over time seemed to be neither feasible nor worthwhile, what agents did over the whole cycle of policies was taken into account.
The creation of a single network that contains all the connections observed in each one of the waves, could be considered the first methodological step in the process of analysis, but with it other considerations become fundamental. In fact, the creation of an aggregate network implied the compression of the time dimension into a single state. This clearly produced the loss of what was an essential element of the case study: the time dimension. The attempt to re-introduce it in the analysis has been one of the major challenges of my research project.

To summarize, two elements emerged after the initial view of the case study (and so after the identification of the first research question). These two elements, which are the specific object of the policies (innovation processes) and the sequence in which these policies were implemented (the waves), determined the elaboration of new and more detailed research questions. These, in turn, immediately guided me in the development of what I consider to be a coherent approach. Since the importance of organizations has been highlighted thanks to David Lane's innovation theory (paragraph 1.2), I wanted the methodologies I was looking for to be aligned as much as possible with the concepts developed. On the other hand, the consideration of a case study in which the essential elements are interactions, necessarily gave a strong importance to the temporal dynamics that occurred and that characterized these interactions. Unfortunately, the specific features of the case study did not allow me to investigate these dynamics by observing the first and most important sequence that characterizes the cycle of policies. The nine observed waves are affected by a high degree of discontinuity regarding agents' participations and partnerships' compositions. This fact forced me to collapse the nine networks (each one representing one of the nine waves) into one (graphs 1.11), and I proceeded with the analysis focusing my attention exclusively on the resulting aggregated network. Even if a compression of the time dimension has been forced, temporal dynamics are introduced through the tailoring of specific models belonging to different methodologies. Before explaining how models were set, the choice of methodologies has to be discussed.

## 1.4. Methodologies

In the description of the theoretical framework in which the analysis was contextualized, specific elements emerged. After having identified the concept of 'organizations', and after having explained its importance with regard to this work, three aspects were introduced: structure, processes and functions[18]. Since these aspects are determinant in the investigation of the object of the whole work, the attempt was to identify three different methodologies that could make possible an appropriate analysis of all of them. Hereunder, how the methodologies were selected is explained.

First of all, an important aspect has to be discussed. There are many methodologies that allow the identification of communities but there is one fact that determines a separation into two distinct groups: there are methodologies that do not make it possible to detect overlapping communities and there are methodologies that make it possible to do this. Since in a context like the one I took into account there are no theoretical reasons that suggest the presence of communities that have to be disjoint, immediately all the methodologies that did not allow me to find overlapping communities were excluded. Moreover, considering the fact that exclusively innovative projects were supported, the identification and the study of agents that belong to more than one community can be seen as an element of interest. Thus, all three methodologies that are here introduced, and that later in the work are better discussed, are characterized by the possibility they have of detecting overlapping communities.

[18] As introduced in paragraph 1.2, the *structure* is the framework around which organizations are built. What an organization is able to implement depends first of all on the nature, the shape and the intensity of the relations in which agents belonging to the organization itself are involved in. The concept of *process* regards the ability of agents to activate and to make work streams arise. This is the second step in the definition of entities that are not only characterized by the presence of a structure, but that are characterized by being able to use their structure to enact actions. Finally, what gives sense to all the activities of an organization is its *function*. The ability to enact processes has to be followed by a common objective. This is a condition that has to be investigated because, without the presence of a shared function, an organization has no direction that characterizes and leads its present activity and its future evolution.

Starting from the first aspect that emerged, the one regarding the concept of *structure*, what seemed to me most coherent was to analyze the relational framework that agents constituted during their participation in the considered public policies. The intent was to identify communities of agents focusing attention on the connections they activated, so what I aimed for was a methodology focused on the analysis of the whole structure determined by the edges. Among the possible methodologies, Clique Percolation Method (CPM) by Palla *et al.* (2005) was selected. First of all, one the most important works regarding community detection methodologies in social network analysis (S. Fortunato 2010) highlights CPM as one of the most popular techniques that makes possible the detection of overlapping communities. For its conceptual simplicity and for having been one of the first methodologies to investigate overlapping communities, the CPM algorithm is used in many works and is often used as a comparison for new proposals. Of course, other algorithms exist. A first class of them uses local functions, typically related to the edge density, to evaluate the presence of a cluster (Baumes *et. al* 2005). In some versions of this class, edges are added and removed continuously until an optimal solution is found (Iterative Scan). In another similar version of this class of algorithms, the most important vertices (the importance of which is usually determined in accordance with some centrality scores) are progressively removed in order to disconnect the graph into small components (RAnk REmoval). Other methodologies optimize quality functions, that can be considered as extension of modularity by Newman (Newman 2004, 2006) and that are based on membership coefficients of nodes, to measure the quality of possible overlapping community structures (Zhang *et al.* 2007), or use information regarding vertices similarity to study the probability that nodes belongs to specific communities (Nepusz *et al.* 2008), or other 'extended' versions (developed to detected overlapping communities) of modularity density by Newman (Shen *et al.* 2009, Chen *et al.* 2014). The reasons why I selected CPM were:

- when I started my collaboration with the working group, CPM analysis was in a state of stasis since, after it had been used in a study aimed at detecting inter-cohesive nodes in some specific waves (Russo and Rossi and Caloffi 2011), its application to the entire policy cycle presented some problematic aspects regarding the selection of the value of *k*. I saw trying to give an answer to these difficulties as an important opportunity to continue this path of analysis;
- since the other two methodologies (Dynamic Cluster Index analysis in particular) required much effort and much attention, the most appropriate choice was to use CPM. In fact CPM, while maintaining perfectly the coherence with the aim of the analysis, is characterized by simple methodological steps, and a specific freeware and well-explained software (CFinder) is available.

For these reasons, I selected CPM as the methodology through which to investigate the presence of communities, in order to focus the attention on the structure determined by agents' connections.

The discussion about the second and the third methodologies is simpler, since no particular alternatives were found to be compatible with what I wanted to analyze. First of all, to study how groups of agents dealt with processes, I needed to apply a methodology that can take into account flows that occurred (or that could have occurred) in the observed network. Infomap by Rosvall *et al.* (2008, 2011, 2014) is a recently developed technique that investigates exactly the presence of communities by observing how simulated flows spread over the network. This methodology is oriented towards the detection of groups of agents on the basis of their ability to manage flows. These flows, which can be interpreted as circulation of information or as work streams, can well represent the concept of *processes* introduced above.

Finally, to detect communities characterized by the presence of a shared *function* - that has to be seen as the presence of a common aim among agents - a completely new methodology that has never been applied before to a socio-economic system was used. Dynamic Cluster Index analysis (DCI) by Villani *et al.* (2013, 2015) is a procedure that studies the presence of functional groups by observing the activities of their agents over time. The idea is that agents that share a common function necessarily have integrated behaviors. Thus, through the observation of what agents did over the period of the considered public policies, subsets are investigated to see whether agents showed particular behavioral patterns that can reveal the presence of a common function, rather than having random behaviors. The potential of this methodology lies in its capacity to detect groups of agents without considering any information regarding the topology of the network (and so without considering the relational structures), but focusing exclusively on what agents did in the considered period. Up to now DCI has mainly been applied in simplified contexts (like small artificial network models). Its initial version (the one that was available at the beginning of this work) was tested on an artificial system of 12 nodes observed in 48 instants over time and so the its introduction in a context where 1.121 agents were involved in only 9 waves over six years, made problematic aspects emerge that the authors had never encountered before. The process of adaptation of DCI procedure to the specific context of this thesis implied several interventions that regarded: (i) the informational basis, (ii) the metric of the time dimension, (iii) the reduction of the number of agents and, lastly but probably most importantly, (iv) the construction of a coherent procedure that could allow me to use this methodology in a context of this kind.

In the central part of the thesis these methodologies are described, and for each one of them some specific settings are discussed. The possibility to introduce and to elaborate particular aspects led to the configuration of different models for each methodology considered. The development of these proposals constitutes the most important element of originality of my work.

# 2. Informational basis

The available dataset that the research group allowed me to use contains unique information. Thanks to this particular informational basis, all the methodologies could be applied. In this chapter I describe its main features and I explain in particular how I managed it in order to elaborate the network database. Elaborations were carried out using STATA/SE.13 and R 3.2.0.

The initial dataset contains information about the 168 network projects funded by the regional government of Tuscany from 2002 to 2008. At the beginning of my work, I merged this database with a specular database regarding the 130 projects that were not funded in the context of the same cycle of policies. The resulting dataset, whose variables are described in table 2.1, contains information about the features of agents (yellow color in table 2.1) which participated in the considered public policies, and it also contains information about the features of the projects carried out (orange color in table 2.1). After the described merge, I checked the values of some important variables (see notes ‘c’ in table 2.1), then I reclassified some variables (see notes ‘r’ in table 2.1) and finally I elaborated new variables that were crucial for the development of the analysis (see notes ‘e’ in table 2.1). In particular, I generated (i) variables that summarize the activity that the agents had over the whole cycle of policies and (ii) variables regarding the profile of the activity of agents over time. Both are highlighted with green color in table 2.1. Even if the database contains information regarding agents’ features and projects’ features, each observation entails a specific combination between an agent and a project (in which the agent participated) and so it has to be considered a database of participations. Information regarding the features of each agent are repeated every time the participation involves that specific agent, and information regarding the features of each project are repeated every time the participation regards that specific project.

While, in order to apply DCI analysis, the information needed was available after the elaboration of variables regarding agents’ profiles of activity over time, the application of CPM and of Infomap required the elaboration of a network dataset. Thanks to the combined use of STATA/SE.13 and of R 3.2.0 (the network dataset was elaborated using ‘igraph' package), I created network data from the database of participations. Since in this database every observation refers to a different participation, it includes the same kind of information as a bipartite network. Every agent is connected exclusively with the projects in which it participated, and this determines the presence of two partite sets of nodes: agents and projects. In this way, the dataset was used to generate a bipartite network (graph 2.3).

However, since the whole analysis regards the interactions that occurred among agents, a bipartite representation cannot satisfy the need to observe connections within each pair of agents. For this reason, I created the corresponding unimodal undirected network of agents: I transformed the affiliation matrix in an adjacency matrix in which every node represents one of the agents involved. Of course, this implied an important assumption: each agent, through its participation in a project, was considered to be connected to all other agents that participated in the same project. From an economic perspective, it is not obvious to assume that the participation (in a project) implies the realization of connections to all other agents

*Table 2.1. Description of the database. Additional information below.*

| Variable | Description | Notes | |
|---|---|---|---|
| **ag_id1121** | Progressive (from 1 to 1.121) Identification number of agents with at least one participation in funded project | ❀ | e |
| **ag_id352** | Progressive (from 1 to 352) Identification number of agents with at least two participations in funded projects | ❀ | e |
| **cf** | Tax code | ❀ | c |
| **p_iva** | VAT number | ❀ | c |
| **ag_name** | Denomination of the agent | ❀ | c |
| **ateco02_2dgt** | Standard industrial classification: Ateco 2002 subsection (2 digits) | ❀ | |
| **ateco02_3dgt** | Standard industrial classification: Ateco 2002 division (3 digits) | ❀ | |
| **dimclass** | Dimensional class (categories, see below) | ❀ | r |
| **sll** | Local labour system code | ❀ | |
| **prov** | Province | ❀ | |
| **typology** | Typology of agents (see below) | ❀ | r |
| **proj_id** | Progressive (from 1 to 168) number of the project | ❀❀ | e |
| **proj_name** | Denomination of the project | ❀❀ | |
| **date_beg** | Beginning date of the project | ❀❀ | |
| **date_end** | Ending date of the project | ❀❀ | |
| **funded** | Dummy: funded project or not | ❀❀ | e |
| **tech_dom** | Technological domain of the project | ❀❀ | r |
| **wave** | Wave in which the project was carried out | ❀❀ | |
| **constrain** | Dummy: presence of constraint in the corresponding wave or not | ❀❀ | e |
| **proj_dim** | Number of agent involved in the project | ❀❀ | e |
| **year** | Year in which the project was carried out | ❀❀ | |
| **n_proj_fund** | Number of participations that the agent had in funded projects | ❀❀❀ | e |
| **n_proj_nofund** | Number of non funded projects proposed by the agent | ❀❀❀ | e |
| **n_prop** | Number of proposals (of projects) made by the agent | ❀❀❀ | e |
| **bool_*** | Series of 59 boolean variables describing the activity of the agents (variable equal to 0 if the agents is active, 0 if not) over the identified instants in time | ❀❀❀ | e |
| **lev_act_*** | Series of 59 variables describing the number of projects in which the agent was participating in each defined instant over time | ❀❀❀ | e |
| **connAg_*** | Series of 59 variables describing the number of agents to which the agent was supposed to be connected through common participations in the projects that were active in each defined instant over time | ❀❀❀ | e |
| **dloa_*** | Series of 58 variables describing the variation (over time) of the number of projects in which the agent was participating (see note below) | ❀❀❀ | e |

❀ = variables describing agents' features ❀❀ = variables describing projects' features ❀❀❀ = variables describing agents' activity

* = series of variables with the same prefix

e = elaborated variable c = checked values r = reclassified variables

Construction of variables dloa_*:
- if the agent has a number of active projects higher than the number of projects the agent was participating in, in the previous instant, the corresponding variable assumes value 3;
- if the agent has a number of active projects equal to the number of projects the agent was participating in, in the previous instant, and the number is greater than zero, the corresponding variable assumes value 2;
- if the agent has a number of active projects lower than the number of projects the agent was participating in, in the previous instant, and the number is greater than zero the corresponding variable assumes value 1;
- if the agent is participating in no funded projects, it has to be considered inactive and so the corresponding variable assumes value 0.

| Typology | | |
|---|---|---|
| **Label** | **Description** | **Notes** |
| **GI** | Government Institutions | |
| **TA** | Trade Associations | |
| **CC** | Chambers of Commerce | |
| **SC** | Service Centers | |
| **LGI** | Local Government Institutions | |
| **E** | Enterprises | |
| **R** | Universities and Public Research Centers | |
| **PRC** | Private Research Enterprises | |
| **BS** | Business Services Enterprises | |
| **KIBS** | Knowledge-Intensitive Business Services | According with The European Monitoring Centre on Change (EMCC), KIBS have been identificated on the basis of available standard industrial classification (ATECO 2002). Agents with one of the followings classifications have been defined as KIBS: 72.1, 72.2, 72.3, 72.4, 72.6, 73.1, 73.2, 74.1, 74.2, 74.3, 74.4 |

*Table 2.2. Reclassification of the variable 'Typology'. KIBS class was introduced.*

involved (in the same project). It is likely that two agents that participate in the same project can easily establish a connection between them, but this is not compulsory[19]. At the same time, it has to be remembered that the nature of the objective of the policies, and so the target of each project, was so specific (all projects regarded the development of innovative processes) that it is feasible to think that all interactions among participants involved in the same project truly occurred. More detailed information (regarding specific connections between agents) was not available and so I proceeded with the transformation in a unimodal undirected network.

The creation of a unimodal undirected network made it possible to elaborate an edge weight structure based on the number of participations in common projects that every couple of agents had. This weight structure was used in the application of the Infomap algorithm, where it was essential from two perspectives: (i) in the simulation of a memoryless flow, I imposed that where the flow moves depends on a probability set determined by the weight of the edges to which the flow can move[20]; (ii) since in all Infomap models the teleportation[21] can allow the flow to move in one step to any edge of the network, I imposed that the probability that edges were the landing point of this 'jump' was calculated on the basis of the edge weight structure.

I conclude with some final points about the handling of two specific aspects regarding the informational basis used for the analyses. First, at the beginning information regarding non funded projects was added and was used for the creation of some important variables (like variable 'n_prop', see table 2.1). However, after some considerations, it was decided to exclude it from the analyses and, what is more important to underline, from the creation of

[19] Two agents can participate in the same project working in different technological areas, or at different phases of the process, without necessarily establishing a one degree connection.

[20] See paragraph 4.3.1

[21] See paragraph 4.3.2

the unimodal network. The whole work aims to investigate the interactions that occurred among agents and thus, even if the submission of a non funded proposal implies an initial interaction/coordination among agents, the activities and the processes that are the true core of the project were not realized. Second, as I explain in paragraph 5.3.3, DCI analysis required a reduction in the number of agents considered for the analysis. Thus, for a coherent development of all the analyses, the set of agents that was used for all three methodologies was always the same: the 352 agents that had at least two participations in funded projects.

*Graph 2.3. Bipartite graph of participations (to funded and to non funded projects) occurred in the context of the entire policy cycle. Aquamarine circular vertices are agents (1.617). Red squared vertices are projects (298). Fruchterman Reingold layout.*

# 3. Clique Percolation Method

## 3.1. Concepts and the problematic definition of k

Clique Percolation Method (CPM) was first introduced by Palla *et al.* (Palla *et al.* 2005). It is one of the most popular methods to detect overlapping communities in a network and, for sure, it is a reference point in literature. CPM is based on the concept of *clique*. A clique is a subset of vertices all connected among them: every vertex is adjacent to every other vertex, and so a clique defines a complete subgraphs, where a complete subgraph is a part of a graph in which exist all possible edges. The dimension of a clique is identified by the number of nodes it contains (usually in graph theory the number of nodes included in a clique is represented with letter *k*) and it follows that the number of edges is always equal to [*k**(*k*-1)/2]. After having investigated the presence of cliques in the observed network, CPM defines communities on the basis of some kinds of extensions of cliques. Fortunato (2010) explains clearly the concepts that are at the core of CPM:

> Two k-cliques are *adjacent* if they share (k − 1) vertices. The union of adjacent k-cliques is called *k-clique chain*. Two k-cliques are *connected* if they are part of a k-clique chain. Finally, a *k-clique community* is the largest connected subgraph obtained by the union of a k-clique and of all k-cliques which are connected to it (…). One could say that a k-clique community is identified by making a k-clique 'roll' over adjacent k-cliques, where rolling means rotating a k-clique about the k−1 vertices it shares with any adjacent k-clique. By construction, k-clique communities can share vertices, so they can be overlapping. (Fortunato, 2010)

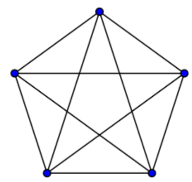


*Figure 3.1. Example of complete graph made up by 5 nodes. All nodes are connected among them.*

Thus, communities detected by CPM are characterized by a high density of connections among nodes. Without calculating any kind of indices, CPM focuses attention exclusively on the structure of edges. For this reason, it is a linear and simple method and in addition it was one of the first methods - and for its simplicity one of the most used - to allow nodes to belong to more than one community. With CPM communities can overlap, and this is a crucial aspect in terms of the aim of this work.
Even if CPM seems very easy to apply and its functioning is not complicated, it is very important to underline the fact that the results obtained with CPM necessarily have to do with an element that can be very difficult to define: the value of *k*. The community structure that is found through the 'rolling' of cliques completely depends on the clique dimension that is used, and so on this value. This aspect is crucial and it can cause some uncertainties, especially if there are no specific theoretical reasons that could help to determine it *a priori*.

The software that allows the application of CPM is CFinder. This software requires exclusively a description of the network (vertices and edges). No weights are required, since they do not have any importance in this analysis[22], and no parameter has to be set. Just with the information regarding the network of the structure, CFinder produces results for any possible value of *k*, from *k* equal to *3* to the maximum that can be reached in the observed network. This means that for every value of *k*, CPM produces a specific partition with its own communities. This situation can cause some problems, since it can be difficult to orientate oneself in such a quantity of results. Thus, the main point concerns the definition of the value of *k*: *which is the clique's dimension that is the most coherent with the analysis that is developed in this work?* Only by answering this question can we determine a specific value of *k* and, consequently, a single partition among all those detected.

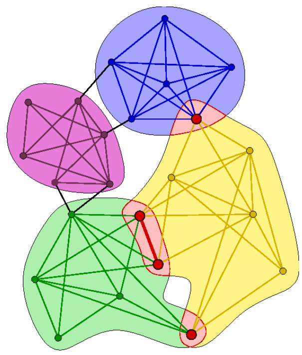


*Figure 3.2. Example of detection of communities with Clique Percolation Method. Communities are detected with k=4*

This question, for the moment, has no answer. The communities that are investigated in this work can have different relational structures and there is no particular reason to establish *a priori* a specific value of *k*. In fact, considering innovative communities, no particular reason could be found to set the value of *k* at any specific value. Of course, this is not a very comforting situation. However, further considerations can be made shifting to a different perspective. In fact, what I have explored in this work, is the evaluation of the values of *k* observing the differences among the different partitions[23]. Instead of trying to find possible motivations that could suggest specific values of *k*, the idea was to observe the features of every single partition and to make a comparison among all of them. Since there are no theoretical guidelines that suggest the assumption of specific values of *k* in a context that concerns communities of innovative agents, what seemed to be the most sensible solution was to look at the structure of the entire partition that every possible value of *k* produced. If this perspective is accepted every value of *k* can be selected, but only few of them need to be chosen since the study of every possible partition would quickly lead to an incomprehensible situation. Thus, what I propose in this work is to analyze the partitions and, on the basis of this observation, to select the most meaningful values of *k*.

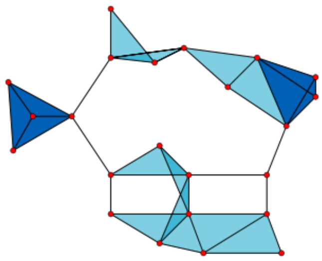


*Figure 3.3. Example of detection of communities with Clique Percolation Method. Different values of k are used. Light blue areas identify communities detected with k=3, while dark blue areas identify communities detected with k=4.*

In the following paragraph the process of selection of values of *k* will be introduced and explained. This operation has to be considered a phase of the setting of this methodology. Even if the normal order has been overturned - because there should be a theory that supports the decision about the value of *k* - there are elements to consider this procedure

[22] CFinder allows the possibility to the user to make a restriction of the network through a selection of nodes that can be made on the basis of their weights. However, from the point of view of the functioning of the method and of the algorithm, weights do not count.

[23] Every partition is determined by a specific value of *k*.

coherent and admissible considering the context of this work. Once again, given that no specific theory suggests specific values of *k*, the decision was to shift the focus immediately to the kind of partition that every single value of *k* produces.

## 3.2. Application of CFinder software

CPM was applied to the subset of 352 agents with at least 2 participations in funded projects. Information about non funded projects were not taken into account and no weighting structure was used because, as explained above, no weighting structure is needed. The algorithm of CPM does not need any particular configuration and so results can be obtained very easily. CPM calculates communities for every possible value[24] of *k,* and produces for every one of it a different partition of the graph. Communities' dimensions, overlaps among communities, number of communities to which nodes belong, number of nodes that are included in the revealed community structure, etc.: these are just some possible elements that can differ among partitions. Thus, it should be clear that different values of *k* produce completely different schemes of the network community structure. Some of them, have to be regarded with interest, while others produce partitions with characteristics that do not seem to be very coherent with the possible organizational scheme that is being investigated.

### 3.2.1. Results of CFinder

Starting with the observation of values regarding the dimension of communities detected with different values of *k*, some considerations can be made. However, no specific information that could suggest particular values of *k* (see chart 3.4 and table 3.5) can be made. The only information that can be regarded with interest is the strong stability in the distribution of communities' dimensions that is achieved with values of *k* higher than 9. It can be observed that the maximum dimension and the minimum dimension[25] of detected communities converge dramatically from *k* equal to 9 onwards. In fact, from that point on, the average dimension of communities stabilizes around the value of 21 and its standard deviation also stabilizes, depicting partitions characterized by homogeneous communities' dimensions. The only relevant difference that can be observed among values of *k* higher than 9, is a progressive decrease in the number of involved nodes, a fact that is normal with respect to the increase in the dimension of the clique considered. In any case, the quick convergence that is observed for $k>9$ deserves attention.
More interesting observations can be made looking at information about communities' overlaps. In fact, observing the chart 3.6, some considerations can be made about the structure of the partition that every value of *k* produces:

---

[24] The possible values of *k* are included in the range between 3 and the dimension of the maximal clique that is present in the graph under analysis.

[25] Obviously also the average dimension and the standard deviation of communities' dimension converge.

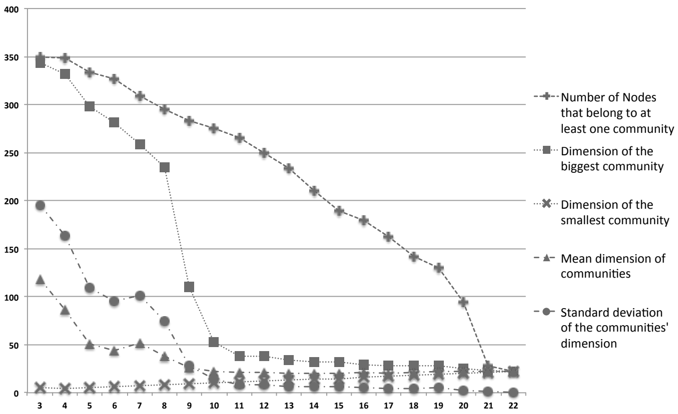



*Chart 3.4 (above). On Y-axis the number of agents, on X-axis the value of k. Different lines represent different statistics that refer to the partition identified using different values of k. These statistics concern the different aspects of the dimension of the communities.*

*Table 3.5 (below). Statistics describing the different partitions detected using different values of k. The statistics are plotted in the chart above, with the exception of the first one (first row) which regards the number of communities detected.*

| K | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Number of communities** | 3 | 13 | 15 | 20 | 24 | 21 | 28 | 32 | 31 | 27 | 24 | 21 | 16 | 17 | 13 | 10 | 8 | 6 | 2 | 1 |
| **Dimension of the biggest community** | 344 | 332 | 298 | 281 | 259 | 235 | 110 | 53 | 38 | 38 | 34 | 32 | 32 | 29 | 28 | 28 | 28 | 25 | 23 | 22 |
| **Dimension of the smallest community** | 5 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 14 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
| **Mean dimension of communities** | 118 | 87 | 51 | 44 | 52 | 38 | 27 | 22 | 21 | 21 | 20 | 20 | 20 | 21 | 21 | 21 | 22 | 22 | 22 | 22 |
| **Standard deviation of the communities' dimension** | 195 | 164 | 109 | 96 | 101 | 74 | 28 | 14 | 8 | 8 | 7 | 6 | 6 | 5 | 4 | 5 | 5 | 3 | 1 | - |
| **Number of Nodes that belong to at least one community** | 350 | 349 | 334 | 327 | 309 | 295 | 283 | 275 | 265 | 250 | 234 | 210 | 189 | 179 | 163 | 142 | 130 | 94 | 28 | 22 |

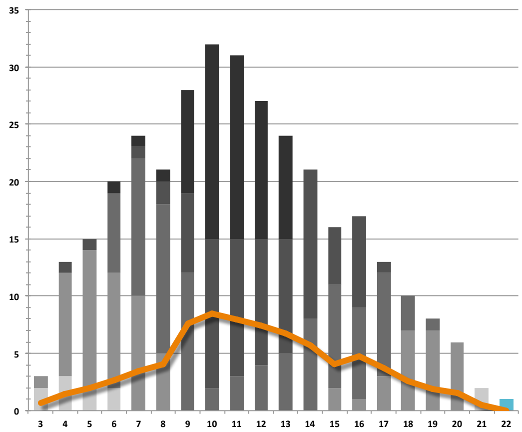



*Chart 3.6 (above). Histogram representing the number of communities detected using different values of k. Bars are divided in different colors representing the number of overlaps that the communities have. The darkest color identifies communities that have more than 15 overlaps with other communities. The lighter is the color the smaller is the number of connections among communities (overlaps). The orange line represents the average number of connections per community. The light blue color identifies communities that do not overlap with any other communities.*

*Table 3.7 (below). Statistics regarding the overlaps among communities that are present in every partition detected using different values of k.*

| *k* | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Number of communities that overlap with more than 15 communities** | | | | 1 | 1 | 1 | 9 | 17 | 16 | 12 | 9 | | | | | | | | | |
| **Number of communities that overlap with a number of communities inclued in the range [11:15]** | | 1 | 1 | | 1 | 2 | 7 | 13 | 12 | 11 | 10 | 13 | 5 | 8 | 1 | | | | | |
| **Number of communities that overlap with a number of communities inclued in the range [6:10]** | | | | 7 | 12 | 13 | 12 | 2 | 3 | 4 | 5 | 8 | 9 | 8 | 9 | 3 | 1 | | | |
| **Number of communities that overlap with a number of communities inclued in the range [2:5]** | 1 | 9 | 12 | 10 | 10 | 5 | | | | | | | 2 | 1 | 3 | 7 | 7 | 6 | | |
| **Number of communities that overlap with 1 community** | 2 | 3 | 2 | 2 | | | | | | | | | | | | | | | 2 | |
| **Number of communities that overlap with NO communities** | | | | | | | | | | | | | | | | | | | | 1 |
| **Number of communities** | 3 | 13 | 15 | 20 | 24 | 21 | 28 | 32 | 31 | 27 | 24 | 21 | 16 | 17 | 13 | 10 | 8 | 6 | 2 | 1 |
| **Number of connections among communities** | 2 | 19 | 30 | 53 | 82 | 86 | 214 | 270 | 246 | 200 | 161 | 119 | 65 | 81 | 48 | 26 | 15 | 9 | 1 | 0 |
| **Average number of connections per community** | *0,7* | *1,5* | *2,0* | *2,7* | *3,4* | *4,1* | *7,6* | *8,4* | *7,9* | *7,4* | *6,7* | *5,7* | *4,1* | *4,8* | *3,7* | *2,6* | *1,9* | *1,5* | *0,5* | *0,0* |

- first of all, there are never communities that do not overlap with other communities, with the exception of the only community that is found for *k*=22. The intense structure of overlaps that can always be observed is likely to be a consequence of how the network was built and of the selection of agents (chapter 2);
- even if partitions always reveal an intense degree of communities' overlaps, from values of *k* between 4 and 8 the situation seems to be less chaotic. As far as *k*=7, the majority of detected communities overlap with a number of communities from 2 to 5. With *k*=8 there are 13 communities that overlap with a number of communities included between 6 and 10, a signal of the fact that the partitions are becoming more entangled;
- from values of *k* between 9 and 13 the situation seems to be very chaotic, since there are always at least 8 communities that overlap with more than 15 communities and since there are always at least 7 communities that overlap with a number of communities included between 6 and 10. In addition, it can be observed that the average number of connections[26] that every community has (orange line in chart 3.6) increases rapidly for *k* higher than 8;
- from values of *k* higher than 13, there is a progressive reduction of the number of communities and of the average number of overlaps per community. Partitions seem to depict a less chaotic situation and less communities with a high number of overlaps are observed.

To conclude, even if there is no specific theoretical information about a correct value of *k*, further considerations about its values were made. Rather than focusing the attention on specific community's features, a different perspective was adopted. Looking at the obtained partitions, with special attention to the overlaps among communities, made it possible to reflect upon the selection of *k*. Always conscious that there is not a specific clique size that is more recommended than another one, the observation of the structure of the partitions determined by different values of *k* permitted the identification of situations not in line with what was expected in this context.

### 3.2.2. Three groups of partitions

Since there is no theoretical reason that recommends specific values of *k* more than others, considerations about the observed partitions were taken into account to go further in the selection of the values of *k*. As observed above, three distinct areas could be detected with respect to the features of the partitions:

- the first group of partitions can be associated with values of *k* between 3 and 8. For these values there is a limited number of communities with non homogeneous dimensions. Excluding *k*=3 that can be considered as an outlier, there is a minimum of 13 communities with *k*=4 and a maximum of 21 communities with *k*=8. Moreover, even if they show a high number of connections among communities, these partitions do not represent the most overlapping situations that were found in this analysis;

[26] Connections among communities are defined as overlaps among them. If two communities have at least one node in common, they are considered to be connected.

- the second group of partitions can be associated with values of *k* between 9 and 16. These partitions are characterized by such a high number of overlaps, that the average number of overlaps per community reaches the value of 8,44 (see table 3.10). The structures of communities that were detected are very entangled and they can only represent with difficulties what is being investigated in this work. In a network made up of 352 agents, having 32 communities with 17 of them that overlap with more than 15 groups, does not seem a plausible representation. Certainly, a situation in which the number of overlaps among communities increases so intensively, cannot be regarded with much interest in terms of a reasonable economic partition. However, there is an element that has to be investigated: the sudden convergence of minimum and maximum dimensions of communities reveals the achievement of a homogeneous condition. Something changes very quickly and for this reason the range of values of *k* between 9 and 14 deserves to be taken into account;
- the third group of partitions can be associated with values of *k* between 17 and 22. In these partitions it is possible to observe that the number of communities and the average number of overlaps per community diminish, reaching values very similar to those observed in the first group. With respect to the first group, the difference that has to be underlined lies in the distribution of the dimension of communities. While in the first group of partitions the standard deviation of the communities' dimensions was around 100 units, in this group the same quantity is never higher than 5. The communities of these partitions have almost always the same dimension, and this constitutes a big difference with the first group detected.

As explained above, even if there are no theoretical references that can help in selecting specific values of *k*, it is unfeasible to consider all possible values of *k*. However, thanks to the grouping of partitions just described, the analysis was able move forward focusing the attention on few representative values of *k*. The identification of these ranges of *k* has the purpose to support the process of selection of some specific values of *k*. Since three of these groups were identified, the idea was to select a single value of *k* for every one of them, in order to assume a specific partition as representative of the whole group (of partitions) it belongs to.

| *k range* | **3-8** | **9-16** | **17-22** |
|---|---|---|---|
| **Number of communities** | max = 24<br>min = 3 | max = 32<br>min = 16 | max = 13<br>min = 1 |
| **Standard deviation of communities' dimension** | max = 195<br>min = 74 | max = 28<br>min = 5 | max = 5<br>min = 1 |
| **Number of overlaps among communities** | max = 86<br>min = 2 | max = 270<br>min = 65 | max = 48<br>min = 0 |

*Table 3.8. Main features of the three ranges of k established. In the cells, that maximum and the minimum values of the statistics considered. These aspects are the most determinant aspects that were taken into account to select the ranges of k.*

*Table 3.9. Descriptive statistics regarding the partition belonging to the first range of k. In orange is highlighted the value that at the end was taken as the representative for the whole range. In yellow, other important values that were used as a comparison to decide the most appropriate value of k.*

**First group of partitions (3 ≤ k ≤ 8)**

| k | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|
| Number of communities | 3 | 13 | 15 | 20 | 24 | 21 |
| Dimension of the smallest community | 5 | 4 | 5 | 6 | 7 | 8 |
| Dimension of the biggest community | 344 | 332 | 298 | 281 | 259 | 235 |
| Average dimension of communities | 118 | 87 | 51 | 44 | 52 | 38 |
| Standard deviation of the communities' dimensions | 195 | 164 | 109 | 96 | 101 | 74 |
| Number of nodes that belong to at least one community | 350 | 349 | 334 | 327 | 309 | 295 |
| Number of connections among communities | 2 | 19 | 30 | 53 | 82 | 86 |
| Average number of connections per community | 0,67 | 1,46 | 2,00 | 2,65 | 3,42 | 4,10 |

| k | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|
| Number of nodes that do not belong to any community | 2 | 3 | 18 | 25 | 43 | 57 |
| Number of nodes that belong to 1 community | 345 | 317 | 275 | 245 | 209 | 193 |
| Number of nodes that belong to 2 communities | 5 | 28 | 51 | 64 | 70 | 66 |
| Number of nodes that belong to 3 communities | | 4 | 7 | 15 | 23 | 23 |
| Number of nodes that belong to 4 communities | | | | 2 | 3 | 8 |
| Number of nodes that belong to 5 communities | | | | | 3 | 2 |
| Number of nodes that belong to 6 communities | | | 1 | 1 | 1 | 1 |

| k | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|
| Number of communities that do not overlap with any communitiy | | | | | | |
| Number of communities that overlap with 1 communtiy | 2 | 3 | 2 | 2 | | |
| Number of communities that overlap with 2 communties | 1 | 5 | 4 | 2 | | |
| Number of communities that overlap with 3 communtiies | | 3 | 3 | 4 | 2 | 1 |
| Number of communities that overlap with 4 communtiies | | | 3 | 2 | 1 | 1 |
| Number of communities that overlap with 5 communtiies | | 1 | 2 | 2 | 7 | 3 |
| Number of communities that overlap with 6 communtiies | | | | 1 | 4 | 2 |
| Number of communities that overlap with 7 communtiies | | | | 4 | 4 | 2 |
| Number of communities that overlap with 8 communtiies | | | | 1 | 3 | 2 |
| Number of communities that overlap with 9 communtiies | | | | 1 | 1 | 6 |
| Number of communities that overlap with 10 communtiies | | | | | | 1 |
| Number of communities that overlap with 11 communtiies | | 1 | | | 1 | 1 |
| Number of communities that overlap with 12 communtiies | | | | | | |
| Number of communities that overlap with 13 communtiies | | | | | | 1 |
| Number of communities that overlap with 14 communtiies | | | 1 | | | |
| Number of communities that overlap with 15 communtiies | | | | | | |
| Number of communities that overlap with 16 communtiies | | | | | | |
| Number of communities that overlap with 17 communtiies | | | | | | |
| Number of communities that overlap with 18 communtiies | | | | | | |
| Number of communities that overlap with 19 communtiies | | | | 1 | | |
| Number of communities that overlap with 20 communtiies | | | | | | 1 |
| Number of communities that overlap with 21 communtiies | | | | | | |
| Number of communities that overlap with 22 communtiies | | | | | | |
| Number of communities that overlap with 23 communtiies | | | | | 1 | |

*Table 3.10. Descriptive statistics regarding the partition belonging to the second range of k. In orange is highlighted the value that at the end was taken as the representative for the whole range. In yellow, other important values that were used as a comparison to decide the most appropriate value of k.*

**Second group of partitions (9 ≤ k ≤ 16)**

| k | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|
| Number of communities | 28 | 32 | 31 | 27 | 24 | 21 | 16 | 17 |
| Dimension of the smallest community | 9 | 10 | 11 | 12 | 13 | 14 | 14 | 16 |
| Dimension of the biggest community | 110 | 53 | 38 | 38 | 34 | 32 | 32 | 29 |
| Average dimension of communities | 27 | 22 | 21 | 21 | 20 | 20 | 20 | 21 |
| Standard deviation of the communities' dimensions | 28 | 14 | 8 | 8 | 7 | 6 | 6 | 5 |
| Number of nodes that belong to at least one community | 283 | 275 | 265 | 250 | 234 | 210 | 189 | 179 |
| Number of connections among communities | 214 | 270 | 246 | 200 | 161 | 119 | 65 | 81 |
| Average number of connections per community | 7,64 | 8,44 | 7,94 | 7,41 | 6,71 | 5,67 | 4,06 | 4,76 |

| k | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|
| Number of nodes that do not belong to any community | 69 | 77 | 87 | 102 | 118 | 142 | 163 | 173 |
| Number of nodes that belong to 1 community | 155 | 131 | 136 | 131 | 123 | 112 | 110 | 99 |
| Number of nodes that belong to 2 communities | 76 | 87 | 70 | 68 | 70 | 52 | 47 | 39 |
| Number of nodes that belong to 3 communities | 29 | 24 | 25 | 24 | 22 | 30 | 25 | 28 |
| Number of nodes that belong to 4 communities | 9 | 18 | 11 | 15 | 10 | 10 | 5 | 7 |
| Number of nodes that belong to 5 communities | 7 | 5 | 13 | 6 | 3 | 3 | 1 | 4 |
| Number of nodes that belong to 6 communities |  | 3 | 4 | 1 | 2 | 2 | 1 |  |
| Number of nodes that belong to 7 communities | 3 | 3 |  | 3 | 3 |  |  | 1 |
| Number of nodes that belong to 8 communities |  |  | 4 |  |  |  |  | 1 |
| Number of nodes that belong to 9 communities | 2 | 2 | 1 | 1 |  | 1 |  |  |
| Number of nodes that belong to 10 communities | 1 | 2 |  |  |  |  |  |  |
| Number of nodes that belong to 11 communities | 1 |  | 1 | 1 | 1 |  |  |  |

| k | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|
| Number of communities that do not overlap with any communitiy |  |  |  |  |  |  |  |  |
| Number of communities that overlap with 1 communtiy |  |  |  |  |  |  |  |  |
| Number of communities that overlap with 2 communtiies |  |  |  |  |  |  |  |  |
| Number of communities that overlap with 3 communtiies |  |  |  |  |  |  |  |  |
| Number of communities that overlap with 4 communtiies |  |  |  |  |  |  | 1 | 1 |
| Number of communities that overlap with 5 communtiies |  |  |  |  |  |  | 1 |  |
| Number of communities that overlap with 6 communtiies |  |  |  |  | 1 | 2 | 3 | 3 |
| Number of communities that overlap with 7 communtiies | 2 | 1 | 1 | 2 | 1 | 1 | 2 |  |
| Number of communities that overlap with 8 communtiies | 2 |  |  | 1 | 2 | 1 | 3 | 1 |
| Number of communities that overlap with 9 communtiies | 1 |  | 1 | 1 |  | 2 |  | 3 |
| Number of communities that overlap with 10 communtiies | 7 | 1 | 1 |  | 1 | 2 | 1 | 1 |
| Number of communities that overlap with 11 communtiies | 2 | 1 | 5 | 2 | 4 | 3 | 5 | 3 |
| Number of communities that overlap with 12 communtiies | 2 | 1 | 3 | 2 |  | 1 |  | 3 |
| Number of communities that overlap with 13 communtiies | 1 | 2 | 1 | 1 | 1 | 2 |  | 2 |
| Number of communities that overlap with 14 communtiies | 1 | 6 | 2 | 3 | 1 | 3 |  |  |
| Number of communities that overlap with 15 communtiies | 1 | 3 | 1 | 3 | 4 | 4 |  |  |
| Number of communities that overlap with 16 communtiies | 2 | 4 | 2 | 2 | 4 |  |  |  |
| Number of communities that overlap with 17 communtiies | 2 |  | 1 | 2 | 4 |  |  |  |
| Number of communities that overlap with 18 communtiies | 2 | 3 | 3 | 3 |  |  |  |  |
| Number of communities that overlap with 19 communtiies | 1 | 2 | 3 | 1 |  |  |  |  |
| Number of communities that overlap with 20 communtiies | 1 | 4 |  | 2 | 1 |  |  |  |
| Number of communities that overlap with 21 communtiies | 1 |  | 3 |  |  |  |  |  |
| Number of communities that overlap with 22 communtiies |  |  | 1 | 2 |  |  |  |  |
| Number of communities that overlap with 23 communtiies |  |  | 1 |  |  |  |  |  |
| Number of communities that overlap with 24 communtiies |  |  | 1 |  |  |  |  |  |
| Number of communities that overlap with 25 communtiies |  | 1 |  |  |  |  |  |  |
| Number of communities that overlap with 27 communtiies |  |  | 1 |  |  |  |  |  |
| Number of communities that overlap with 28 communtiies |  | 3 |  |  |  |  |  |  |

*Table 3.11. Descriptive statistics regarding the partition belonging to the third range of k. In orange is highlighted the value that at the end was taken as the representative for the whole range. In yellow, other important values that were used as a comparison to decide the most appropriate value of k.*

**Third group of partitions (17 ≤ k ≤ 22)**

| k | 17 | 18 | 19 | 20 | 21 | 22 |
|---|---|---|---|---|---|---|
| Number of communities | 13 | 10 | 8 | 6 | 2 | 1 |
| Dimension of the smallest community | 17 | 18 | 19 | 20 | 21 | 22 |
| Dimension of the biggest community | 28 | 28 | 28 | 25 | 23 | 22 |
| Average dimension of communities | 21 | 21 | 22 | 22 | 22 | 22 |
| Standard deviation of the communities' dimensions | 4 | 5 | 5 | 3 | 1 | |
| Number of nodes that belong to at least one community | 163 | 142 | 130 | 94 | 28 | 22 |
| Number of connections among communities | 48 | 26 | 15 | 9 | 1 | |
| Average number of connections per community | 3,69 | 2,60 | 1,88 | 1,50 | 0,50 | 0,00 |

| k | 17 | 18 | 19 | 20 | 21 | 22 |
|---|---|---|---|---|---|---|
| Number of nodes that do not belong to any community | 189 | 210 | 222 | 258 | 324 | 330 |
| Number of nodes that belong to 1 community | 103 | 97 | 103 | 70 | 12 | 22 |
| Number of nodes that belong to 2 communities | 35 | 33 | 19 | 17 | 16 | |
| Number of nodes that belong to 3 communities | 23 | 10 | 8 | 6 | | |
| Number of nodes that belong to 4 communities | | 1 | | 1 | | |
| Number of nodes that belong to 5 communities | 1 | 1 | | | | |
| Number of nodes that belong to 6 communities | | | | | | |
| Number of nodes that belong to 7 communities | | | | | | |
| Number of nodes that belong to 8 communities | 1 | | | | | |

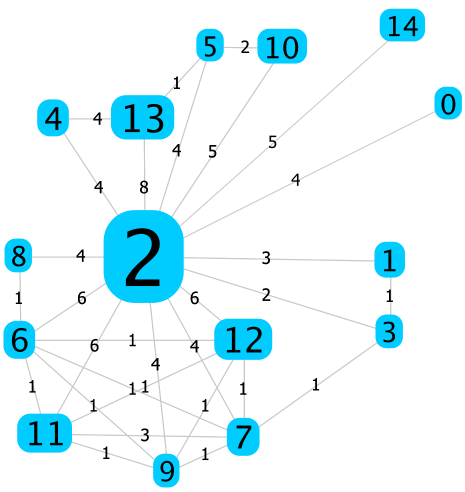
| k | 17 | 18 | 19 | 20 | 21 | 22 |
|---|---|---|---|---|---|---|
| Number of communities that do not overlap with any communitiy | | | | | | 1 |
| Number of communities that overlap with 1 communtiy | | | | | 2 | |
| Number of communities that overlap with 2 communtiies | | | | 2 | | |
| Number of communities that overlap with 3 communtiies | 1 | 1 | 4 | 3 | | |
| Number of communities that overlap with 4 communtiies | 1 | 2 | 3 | | | |
| Number of communities that overlap with 5 communtiies | 1 | 4 | | 1 | | |
| Number of communities that overlap with 6 communtiies | 1 | 1 | 1 | | | |
| Number of communities that overlap with 7 communtiies | 2 | 1 | | | | |
| Number of communities that overlap with 8 communtiies | 3 | 1 | | | | |
| Number of communities that overlap with 9 communtiies | 1 | | | | | |
| Number of communities that overlap with 10 communtiies | 2 | | | | | |
| Number of communities that overlap with 11 communtiies | 1 | | | | | |

## 3.3. Three models of analysis with CPM

After having identified three groups of partitions that correspond to three different ranges of *k*, the idea was to select a single value of *k* for each one of them, in order to assume the correspondent partition as the representative of the whole group. This selection was made through the observation of some more detailed information that helped to understand which value of *k* can be considered the most appropriate in every group. At the end, the three selected values were used to determine three different models of CPM analysis. Every one of these models produces a particular partition that will be examined later. Once again, it is important to remember that none of these possible community structures cannot be considered as wrong. Since no specific theoretical element suggests which is the adequate value of *k* in a context of this kind, all of them are admissible. At the same time, the process of selection of a reduced number of models is necessary to continue the analysis without an excessive proliferation of results. That is the reason why, after having identified three ranges of *k* that correspond to partitions that have been grouped because of their characteristics, only three values of *k* were selected. Each one of them determines a specific model of CPM analysis.

### 3.3.1. k=5 as representative of the first range of k [3; 8]: model CPM_k05

Looking at the partitions that were obtained with values of *k* between 3 and 8, the intent was to choose a value which would make it possible to not have too few communities and which at the same time would not determine a situation with too many connections among communities[27]. Thus, having excluded *k*=3 because it produces a partition with only 3 communities and having excluded *k*=7 and *k*=8 because they correspond to partitions with a too high a number of connections among communities, *k*=4 and *k*=5 and *k*=6 were taken into account. Looking at the number of communities' overlaps, all these three values appear reasonable. In the end, two elements in particular led to the selection of *k*=5. First of all,



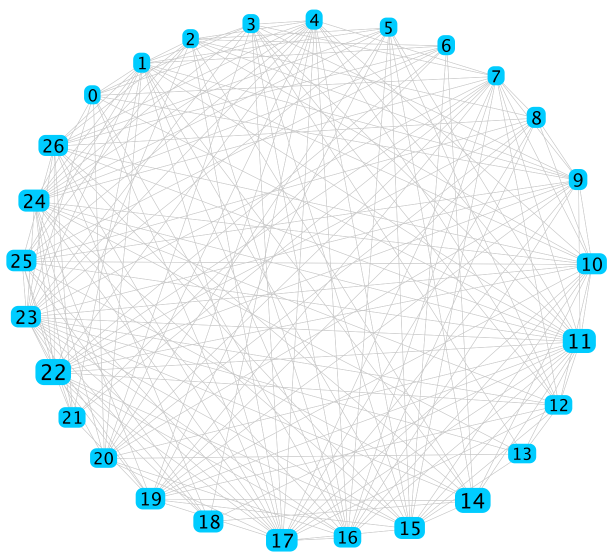
*Graph 3.12. Graph of the partition detected using k=5. This representation was made using CFinder. Every node is one of the detected communities. Edges represent overlaps between communities. The bigger is the node the more agents it includes. However nodes are not proportional to the number of agents included, they are only ranked on the basis of their dimension. The number inside the node represents the number of the community, while the numbers on the edges represent the number of nodes that are present in the corresponding intersection.*

[27] Connections among communities are determined by the presence of nodes in common. If two communities share at least one node, they are connected. Connected communities are overlapping communities.

moving from $k$=5 to $k$=6 a large increase in the number of communities and in the number of connections among them can be observed (+5 communities and +23 overlaps among communities). This suggested focusing attention on $k$=4 and $k$=5, but it was immediately noticed that with $k$=4 there is a single community that includes 94% of agents (332 agents). Moreover, the standard deviation of communities' dimensions diminishes by more than 50 units when moving from $k$=4 to $k$=5. In the end, after these considerations the most adequate choice seemed to be the partition generated with $k$=5. The model associated with $k$=5 was named CPM_k05.

### 3.3.2. k=12 as representative of the second range of k [9; 16]: model CPM_k12

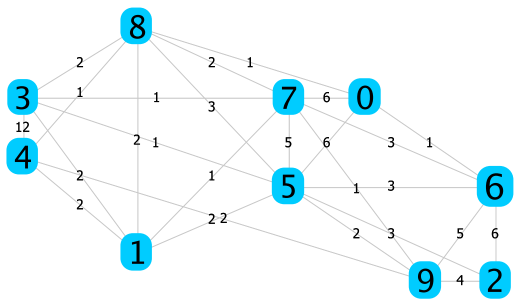
Considering partitions obtained with values of $k$ between 9 and 16, the intent was to investigate the strong and intense reduction in the standard deviation of the communities' dimensions that could be observed in this range of $k$. Of course, the choice of $k$=9 or $k$=10 or $k$=11 would have led to one of the partitions characterized by the highest observed values in terms of number of communities and in terms of overlaps among communities. Even if all the partitions observed for the range of $k$ from 9 to 16 are characterized by a high number of connections, it was not seen as a good solution to select one the values $k$=9 or $k$=10 or $k$=11. The intent was to capture a complex weave of communities but in an attempt to move in the direction of a more dissipated partition. To maintain these two aspects, the value that seemed to be the most adequate was $k$=12. In fact, moving to $k$=13 or to $k$=14 would have caused the loss of more than 20% of connections among communities. It was thought that a partition with a high homogeneity in the communities' dimensions and a low degree of overlaps would be observed in the third identified range of $k$. For these reasons, $k$=12 was selected. The model associated with $k$=12 was named CPM_k12.




*Graph 3.13. Graph of the partition detected using $k$=12. This representation was made using CFinder. Every node is one of the detected communities. Edges represent overlaps between communities. The bigger is the node the more agents it includes. However nodes are not proportional to the number of agents included, they are only ranked on the basis of their dimension. The number inside the node represents the number of the community, while the numbers on the edges represent the number of nodes that are present in the corresponding intersection.*

### 3.3.3. k=18 as representative of the third range of k [17; 22]: model CPM_k18

With the last group of partitions, a very interesting situation was faced: communities are homogeneous in terms of their dimensions and they overlap less than with lower levels of *k*. The problem for these partitions is that the higher *k* is, the fewer agents are included in at least one community. Taking values of *k* higher than 19 would have brought about the inclusion (in the partition) of a maximum of 94 agents, a number that seemed to be too little especially with respect to the fact that an intense selection had already been made at the beginning of the whole analysis (chapter 2). Thus, the focus moved to *k*=17, *k*=18 and *k*=19. For *k*=17 too high a number of overlaps was still observed: there are 4 communities that are connected with at least 9 communities. Since a structure of highly connected communities was already detected for the previous range of *k*, here the idea was to try to observe a more clear and comprehensible partition. For this reason, *k*=17 was excluded.
With regard to *k*=19, it seemed to produce a partition that includes too few agents (130). Moreover, there are few agents that belong to more than one community: there are only 19 agents that belong to 2 communities and only 8 agents that belong to 3 communities. From this perspective, it seemed to be too dissipated a partition and so, in the end *k*=18 was selected. The model associated with *k*=18 was named CPM_k18.



*Graph 3.14. Graph of the partition detected using k=18. This representation was made using CFinder. Every node is one of the detected communities. Edges represent overlaps between communities. The bigger is the node the more agents it includes. However nodes are not proportional to the number of agents included, they are only ranked on the basis of their dimension. The number inside the node represents the number of the community, while the numbers on the edges represent the number of nodes that are present in the corresponding intersection.*

# 4. Infomap

## 4.1. Simulating flows over networks

The second model of analysis that is taken into account in this work, is a methodology called Infomap, whose major contributors are Martin Rosvall and Carl T. Bergstrom[28]. The method they have developed seemed to be very appropriate to be applied to the case study of this work. In particular, since it concerns the detection of communities through the simulation of spreads of flows into a network structure, it can be considered appropriate to investigate how communities could be structured with respect of information circulation among agents. In this chapter the methodological aspects of Infomap are described and how the method was fitted to the specific case of this work will be discussed.

### 4.1.1. Concepts of Infomap

The idea that Rosvall and Bergstrom had, is both very simple and very sophisticated. What they set out to do was to develop process of partitioning a network into communities as a geographical mapping process in which the main issue is to define cities' boundaries. From the perspective of Rosvall and Bergstrom, every community is seen as a city and the problem they pose is: *looking at a map without any kind of political information, how is it possible to assess the boundaries of a city?*

To answer this question, they started to observe real networks, focusing in particular on the structure of the edges and they tried to investigate how its shape affects circulation through it. For example, in a relational network they observed the circulation of information, while in street maps they observed the circulation of means of transport. The crucial point of their methodology lies in the fact that they thought that the process of detection of communities had something to do with the description of these flows that move through the network. To better explain their intuition and to better introduce the methodology they developed, a short example is described hereunder.

### 4.1.2. An explanatory example: street names

If someone moves from the city where they live to a city nearby and looks around to observe the names of the streets, it can be easily supposed that in a short time they will find the same names that they can find in their own city. This image, that seems to be very commonplace, holds the core of the intuition that Rosvall and Bergstrom had. Of course, inside the boundaries of a city it will never happen that two or more streets have the same name but between different cities, street names repeat themselves. This fact aims to simplify the description of places, otherwise if every street in the world has a unique name, longer and longer names will have to be used and in the end nothing will be comprehensible. What in the real world happens is that cities have street names that are often the same and what

[28] Martin Rosvall is an Associate Professor of the University of Umeå (Sweeden), Department of Physics. Carl T. Bergstrom is a Professor of the University of Washington (WA, USA), Department of Biology.

necessarily contributes to distinguish between two streets with the same name but situated in different towns, obviously are the names of the cities. At the same time, if too many cities are defined, after a while the names used to describe them will start to become longer, and this will cause the same effect of inefficiency explained above for street names. Thus, the desirable situation is achieved by taking into account both levels: the number of cities and the number of streets.
To return to the problem of studying networks' partitions, it can be said that the idea of Infomap is to find the community partition that allows this most efficient two-level[29] description of any kind of flows in the network under analysis. In other words, the partition is not studied from the perspective of links that connect nodes, but from the perspective of an efficient description of a simulated flow in the network. The community depicts an area where once a flow has entered, it stays there in a stable and continuous way and once it exits it is not expected to re-enter immediately. Infomap defines communities that detect nodes among which communications are fluent: flows among communities happen, but the majority of flows occur within the communities.
Rosvall and Bergstrom's method relies upon a two-level optimization algorithm that includes two different phases. The first one uses the simulation of a random walk trough the network in order to obtain the description of nodes based on the observation of the frequencies with which nodes are crossed by the random walks. In the second phase, the two-level description of nodes that allows the most efficient description of a random walk is investigated. The first phase relies in particular on a procedure called 'Huffman coding', while the second contains the biggest contribution of Rosvall and Bergstrom: the Map Equation.

### 4.1.3. Huffman coding and the Map Equation

The Huffman coding is an algorithm that provides a unique binary code for elements that are characterized by occurrences. It is a 'prefix-free'[30] code that allows an optimal process of encoding symbols separately. In other words, it is the most efficient (shortest) binary description that can be used to encode sequences of elements. Taking into account the frequencies of occurrences of each element, Huffman coding assigns the shortest codes to elements that occur more frequently, while the fewer elements occur the longer their description is.
With respect to the problem of detecting community in a network, this procedure is important to assign a unique code to every observed node and, more specifically, to assign the shortest descriptions to those nodes that are crossed most frequently by the random walk. Describing a flow over a network as the sequence of the nodes that it touches, Huffman coding makes it possible to minimize the length of the code that describes the flow and it guarantees that nodes have unique names.

---

[29] Referring to the example of street names, the first level of description is associated with cities, while the second level of description is associated with streets' names.

[30] In a prefix-free code the bit string representing some particular symbol is never a prefix of the bit string representing any other symbol. Huffman coding uses a specific method for choosing the representation for each symbol. Huffman coding is the most efficient binary code in the sense that it produces the minimum expected codeword length.

A further reduction of the description length of the considered random walk is impossible: Huffman coding already provides the most efficient solution. This is true, but always under the condition that nodes must have single names. In fact, after a relaxation of this constraint, more efficient solutions can be found and this is exactly what Rosvall and Bergstrom proposed with their Map Equation. Instead of describing the network as a domain where nodes' names cannot be replicated, they introduced a two-level description that produces repetitions in first level nodes' codes but that ensures a unique description of nodes through the introduction of a second level code[31]: a second level that concerns networks' meso-structures like communities. In other words, agents' first level codes could repeat themselves but to every agent is given also a second level code that represents the community it belongs to: the combination of these two codes guarantees a unique code for every node.
At this point, the problem is always the same: to find the way to obtain the most efficient description of a flow that crosses the network. Obviously, while before it was a single level problem, now the problem is a question of the second level description: a partition of the network in communities. The identification of the second level partition that produces the most efficient two-level description passes through a problem of minimization: the partition has to be found that makes it possible to reach the shortest length that describes the random walk. To do this it would be necessary to analyze all possible partitions. Obviously, an exhaustive investigation cannot be carried out because of time computation problem. Thus, the solution is investigated through the application of a deterministic greedy search algorithm[32] at the end of which a specific partition of the network is obtained. At this point, every node has:

- a first level code that is unique in the context of the community it belongs to, but that it can be repeated in the context of the whole network;
- and a second level code that corresponds to the community it belongs to.

As already explained, this two-level description optimizes the length of the description of a random walk that passes through the network. What is more important to underline, is that this methodology leads to the identification of groups that maximize the time over which the simulated flow remains within their agents (persistence time). Communities that are detected applying Infomap algorithm are communities whose main characteristic is to be able to handle almost autonomously processes in which their agents are involved.

---

[31] To explain it easily, the example of the cities can be taken up. Taking for example two cities like Modena and Bologna: if all streets of both towns have to have a unique name, it would be necessary to introduce longer names to streets in order to avoid repetitions. Otherwise, names can be repeated but to describe a specific street it is necessary to specify which city is being referred to. 'Viale Indipendenza' does not indicate a specific place: it should be followed by the name of the city it refers to.

[32] The application of a deterministic greedy search algorithm is followed by the refining of results with a simulated annealing approach using the heat-bath algorithm (Clauset *et al.* 2004, Wakita *et al.* 2007)

## 4.2. Introduction of Markov second order models

As described above, the central element of Infomap algorithm is represented by the propagation of a random walk through the considered network. The simulation that is performed can be set with different parameters that can be managed to reproduce as well as possible the real situation that is observed. Many of these parameters will be introduced later and for the moment attention is focused on the parameter that in the context of this work arouses most interest.

The minimum information that Infomap needs to be launched concerns the network structure. The description of vertices and edges makes it possible to draw a complete picture of the network and, thanks to this, random walks propagations can be studied in order to detect a partition of nodes. Infomap can be loaded with more information about the network and, for example, if weights are assigned to edges or to vertices, the algorithm will elaborate probabilities[33] that will influence the direction of the flow at every step.

Among all possible information that Infomap can use, what reveals its strongest potentiality is that it can take into account what has happened in the real observed network to create the simulated flow. This means that, if information regarding flows that truly crossed the network is available, the algorithm can consider it to produce a simulation that reflects real observed processes. If this information is available, it is possible to create models with a Markov order higher than one. Since the real propagations' dynamics are relevant for the detection of communities, the possibility to apply a models with a Markov order higher than one allows the analysis to become much more interesting.

### 4.2.1. Memory nodes and trigrams

The introduction of models with a Markov order higher than one, means that the direction that the flow will take depends not only on where it is, but also on where it was. In this context, the point that characterizes a Markov model of order higher than one, is that at every step the flow's direction is determined not only by which the vertex that holds the flow at the present stage, but also by the one it was coming from.

In 2014 Rosvall *et al.* introduced this extension to the original form of their Infomap algorithm and the results they obtained were interesting (see Rosvall *et al.* 2014). They took into consideration air traffic among airports to investigate the presence of groups of cities that share intense flows of passengers and, using the Airline Origin and Destination Survey (DB1B) made by the Research and Innovative Technology Administration (RITA), they set up two different models to run Infomap. In the first model, they provided a simple description of the network with nodes' weights depending on the number of passengers that travelled through the airports (Markov order 1). In the second models, after having observed

[33] When the random walk reaches a vertex (or an edge depending on how the analysis is set) it has a specific probability for every adjacent vertex, to move to one of them. If no weights are assigned, every adjacent vertices has equal probability to be reached by the random walk.

every single itinerary[34], they introduced and used information regarding where passengers were coming from (Markov order 2). What they found was that the partition in modules that was provided by Infomap in the second model, was completely different from the first one, and it should be considered more meaningful. Instead of depicting a confused partition as was the case in the first model, in the second model the presence of intense paths of out-and-back from the same cities strongly emerged, and this is meaningful with regard to usual flying itineraries.

The theoretical elements that make it possible to handle this increase of information, are memory nodes. Memory nodes were firstly introduced by Shannon (1948) to allow the development of higher-order Markov models. The basic idea is that every node is characterized by a vector whose elements represent adjacent nodes. In this way when a flow moves from node *i* to node *j*, it does not generically reach node *j* but it reaches memory node *ij*. In other words, memory nodes hold not only the information about the transition of the flow through a node, but they include also the information about where the flow was before coming to that node. It is extremely simple, and at the same time it is extremely crucial. Thanks to this theoretical element, 'memory' has been introduced in models and this means that where the flows move can also depend on where they come from.

Clearly, to implement a higher-order Markov model, as I propose in this work, it is necessary to provide information about memory nodes. With reference to Infomap this can be done with the construction of trigrams. A trigram is a sequence of three nodes that expresses how the flow moved in an observed situation, and they are the only elements thanks to which information about memory nodes can be provided to Infomap software. Obviously, trigrams have to be built, and since no specific application was detected, I implemented a specific code that extracts trigrams, starting from a database of participations in projects. The hypotheses that have to be made to describe flows in the specific case study of this work, are explained in the following paragraphs.

### 4.2.2. Reconstructing flows of information in the case study

Finer implementations of the model are made possible thanks to the application of a second-order Markov condition. The idea that at every step the direction of the flow depends not only on where the flow is but also on where the flow was allows the algorithm to reveal more interesting results. This can be done on condition that data are available since, if no flow is observed in the network under analysis, no second-order Markov condition can be applied.

From this point of view, this specific case study represents a somewhat odd situation. Data about observed innovative flows were not available, but at the same time there was information about the dates of the beginning and of the end of every single project. Even if there was no information about real flows, I set out to consider the possibility to introduce a second-order Markov model using the presence of specific temporal references regarding

[34] A single itinerary could include more than one flight. Rosvall *et al.* (2014) defined the presence of an itinerary, instead of a two or more distinct flight, depending on the booking of the flight plan. If, since the first moment the flight plan is booked, it is included also the return flight, then it as considered a single itinerary. Otherwise, the ticket for return flight is bought in a distinct moment, then it considered to have no continuity and so two distinct itineraries are identified.

the realization of projects. More specifically, I set out to restrict the movement of the simulated flow, forcing it to respect the observed temporal order in the realization of projects. In this way, thinking of how to handle the introduction of a second-order Markov condition, the idea was to construct a fake flow with regard to a number of hypotheses that try to respect what happened in reality. These hypotheses are:

- the flow moved among all agents involved in the same project;
- the flow moved from agents that participated in one project to agents that participated in another project, if there was at least one agent that participated in both;
- the flow moved from agents that participated in one project to agents that participated in another project, if the second project was subsequent and/or simultaneous.

The idea was to exclude the movement of the flow going in the direction of projects that had terminated before the project where the flow came from. It was hypothesized that the flow moved from the participants of a project, to participants of another project, through those agents that participated in both. However, a constraint was imposed that if not respected excludes the presence of the flow: the project where the flow is moving to cannot have finished before the project where the flow comes from has started. It is intuitive to think that if an agent participates in time 1 in a project and in time 2 in another project, what it learned in project 1 could be transferred and used in the context of project 2, while the contrary cannot happen.
An example of how flow dynamics have been constructed using trigrams is explained hereunder. In the example, *x1 x2 x3* and *x4* are agents that participate in network projects of dimension two (every project is characterized by having two participants). Three projects are considered (*A*, *B* and *C*) and each one has a beginning date and an ending date. All events are reported chronologically:

30 Jun. 2014: x1 and x2 start to participate together in project A
01 Aug. 2014: x1 and x3 start to participate together in project B
31 Dec. 2014: x1 and x2 end to participate together in project A
31 Jan. 2015: x1 and x3 end to participate together in project B
01 Mar. 2015: x1 and x4 start to participate together in project C
30 Jun. 2015: x1 and x4 end to participate together in project C

This is exactly the kind of information that is dealt with in this work: participants and projects, with beginning and ending dates, but no specific observations about flows of information or about other kinds of processes. In this context, the approach was to treat all possible flows that respected the conditions explained before as if they had really occurred. With reference to the example, this means that the reconstructed flows would be:

a. flows that move from terminated projects to successive projects
  - x2 participated with x1 in project A, and then x1 participated with x4 in project C:
    x2 - x1 - x4
  - x3 participated with x1 in project B, and then x1 participated with x4 in project C:
    x3 - x1 - x4

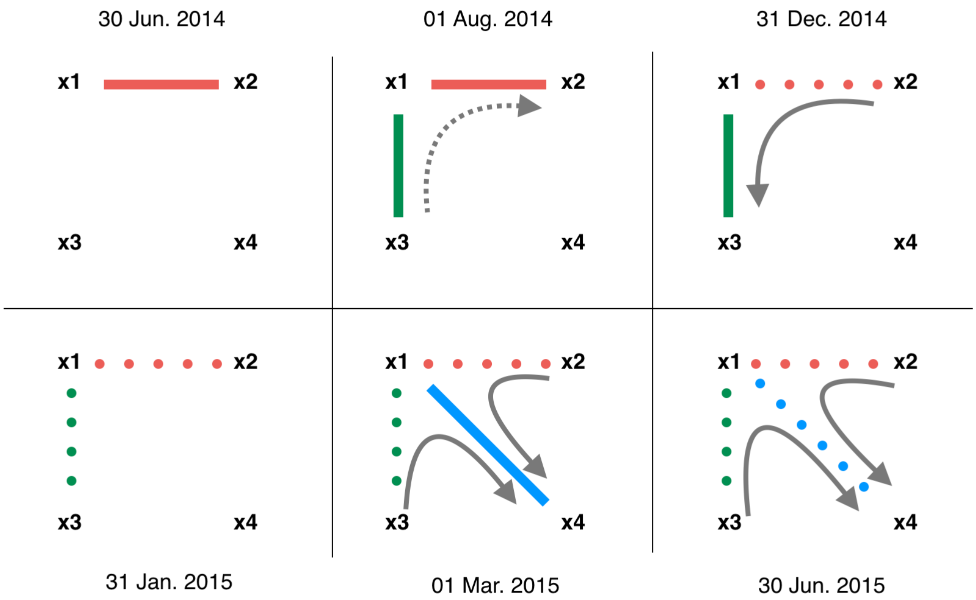



*Figure 4.1. Example representing the situation faced in the case study of this work. Colored lines represent projects (red line represents project A, green line project B and blue line project C). When they are solid the project is active, while when they are dotted the project has terminated. Agents are represented by nodes x1, x2, x3 and x4. Starting from the top left and moving by row until the bottom right, six different instant in time are observed. Grey arrows represent the fake flows that were supposed to have existed. Solid arrows represent flows involving agents that have participated in successive projects. The dotted arrow represents a flow involving agents that are participating in simultaneous projects. Different hypotheses can be made about the admissibility of flows with regard of the temporal perspective.*

• x2 participated with x1 in project A, while x1 was participating with x3 in project B:

x2 - x1 - x3

b. flows that move in the context of simultaneous projects

•x3 participated with x1 in project B, while x1 was participating with x2 in project A:

x3 - x1 - x2

### 4.2.3. Managing flows with respect to simultaneity among projects

Now it can be better understood what was done: the construction of a series of fake flows that could have happened and that respect a simple principle of chronological order. What is interesting to notice is that, due to the fact that projects have a duration in time, a situation of simultaneity can be encountered. The temporal overlap among projects can be handled in different ways that lead to different constructions of trigrams.

Two approaches have been developed in this work. The first approach establishes the presence of a flow when the beginning date of the project where the flow comes from, is anterior to the beginning date of the project where the flow can move to. The means that a flow of information can move from the context of a project immediately after the project has begun. The point is this: to be able to transfer information from a project, agents simply have to start their participation in it. Following this first approach, trigrams that describe the example above are:

x2 - x1 - x4

x3 - x1 - x4

x2 - x1 - x3

x3 - x1 - x2

The second approach considers the presence of a flow only when the ending date of the project where the flow comes from, is anterior to the beginning date of the project where the flow can move to. This means that a flow of information can move from the context of a project, only when the project has terminated. The point is this: before being able to transfer information accumulated in a project, agents have to finish their participation in it. Following this second approach, trigrams that describe the example above are:

x2 - x1 - x4

x3 - x1 - x4

x2 - x1 - x3

In this second approach, the trigram 'x3 - x1 - x2' is not included because project B terminated after project A terminated. For a part of the period in which they were active, project A and project B were simultaneous, but project B terminated only after project A had terminated. Following the interpretation of this second approach, this prevented the presence of a flow from project B to project A.

To conclude this paragraph it is important to underline two further aspects related to the construction of trigrams. Firstly, the choice was made to exclude the presence of 'self-links'. Theoretically a flow could remain for one period (or more) in a node without moving to adjacent nodes. This situation can be represented by a trigram of this kind

x2 - x1 - x1

Since it was not possible to observe real flows through the network and since the decision was to reconstruct a kind of fake flow, in every node a self-link would have been admissible. This would have no particular sense with respect to this specific case study and so, to simplify the analysis, the final decision was to exclude self-links[35].

The second aspect regards the possibility that a flow, after a single step, could immediately come back to the node where it was. Since the intention was to detect communities, this possibility was excluded. If the flow can make a sort of out-and-back itinerary in two steps,

---

[35] To be formally correct, self-link do not regard only the persistence of the flow in a node. They are also used, at the beginning of some trigrams, to indicate that the flow is starting from that point. Thus, in the case there are three agents z1 z2 and z3, if the flow starts from z1 and then moves to z2 and then moves to z3, the trigrams that describe this situation are:

z1 - z1 - z2

z1 - z2 - z3

As it can be observed, in the first trigram there is a self-link, but putting a self-link at the beginning of a trigram does not indicate the persistence of the flow in a node. It indicates that the flow started from that node. This fact is very important in the construction of trigram since the presence of self-links at the begging of trigrams indicates that involved nodes have the capability to give rise to a flow.

of course the probability that it remains for a longer time in a limited neighborhood increases. The intention was to explore carefully the boundaries of communities rather than small part of them. This is the reason why the spreading of the flow was encouraged, trying to limit its persistence in small areas. Of course, even if this decision is coherent with the aim of the analysis, it nevertheless does not prevent the flow returning in the same node it came from after two steps. For example, considering three agents *y1 y2* and *y3*, the trigram '*y1 - y2 - y2*' is not allowed but the sequence of these two trigrams '*y1 - y2 - y3*' and '*y2 - y3 - y1*' is allowed. They describe a single flow[36] that, moving from *y1* passes through *y2* to get to *y3* and, after being there, it goes back to *y1*.

## 4.3. Setting other Infomap parameters

After having described the most important characteristics of the Infomap methodology, some other features that regard the functioning of the algorithm have to be introduced. Considerations about them have been made to set the algorithm coherently.

### 4.3.1. Overlapping communities and network weights

In the first version in which it was implemented, Infomap could detect only disjointed communities. Later, in 2011 (Esquivel A. V. and Rosvall M., 2011) the algorithm was implemented to allow also the identification of overlapping groups of nodes. This was possible thanks to the addition of a new step in the algorithm. After the first standard[37] Infomap procedure, in the new phase every node that stands at a boundary among communities is assigned to the communities to which it is close. Then, the description of the flow is recalculated: when the flow arrives at a node, only one of the communities it belongs

---

[36] To describe a single flow with trigrams, it is required to keep the last part of a trigram and to use it as the beginning part of the next trigram. So, considering node y1 y2 and y3, the flow that is described with this sequence 'y1 - y2 - y3 - y1' is represented by these two trigrams:

y1 - y2 - y3
y2 - y3 - y1

[37] 'Standard' in the sense that produces disjoint communities. In the first stage, the algorithm produces a partition with disjoint communities, as the first version of Infomap allowed to to.

to is selected[38]. Then, a heuristic procedure[39] is used to select some of the best possible partitions and, finally, the best one is determined in accordance with the minimization of the description of the flow[40]. Since in the context of this work there are no theoretical elements that suggest the investigation of disjointed communities, in all developed Infomap models the option that allows the detection of overlapping communities[41] was always used.

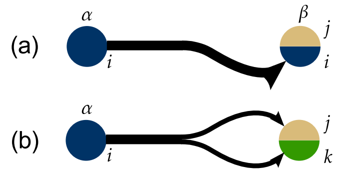



*Figure 4.2. Example of functioning of the procedure of re-calculation of the flow description in a context where overlapping communities are allowed. The re-calculation of the description of the flow, that is necessary to provide the final allocation of nodes to one ore more communities, is made through a new description in which every node is assigned to one of the communities to which it can belong. The flow comes from a node (node α) that was attributed to a specific community (the 'blue' community in the figure) and approaches to a node that still has not received a specific attribution (node β). However, the communities to which node β can belong are already defined. If, as in situation (a), one of these community is the same community to which was attributed node α ('blue' community), then node β is assigned to this specific community. In the case node β can be assigned only to communities that are different to the one to which node α is attributed, then the community to which it will belong will be established on the basis of the available probability set.*

Concerning the importance of weights in Infomap, this has only been mentioned briefly. Infomap in fact allows the introduction of weights and, if they are used, they influence the probabilities (of directions) that the random walk encounters at every step. The mechanism is simple: depending on where the flow is, its probabilities of moving to one of the adjacent nodes (or one of the adjacent edges) is calculated on the basis of the weight of each of these nodes (or edges). Obviously, weights can be determined in many different ways and so considerations regarding how this aspect was managed must be made.

First of all, the intention was to create a structure of weights that can represent the activity that agents had in the context of the regional public policies under analysis. Among possible criteria[42], the one that seemed to be the most appropriate concerned the number of funded

---

[38] It must be noted the selection of the community that occurs when the flow arrives on a node that belong to more than one of them, depends also on the community of origin. As described in figure 4.2, if the flow comes from the community *i* and if the arrival node belongs to community *i* (even if it belongs also to other communities) then the flow remains in community *i*. Otherwise, if the arrival node does not belong to community *i*, one of other communities is selected.

[39] "In the second step, we combine a fraction of all local changes generated in the first step into a global solution. Every time two or more multiply assigned nodes are connected, we need to solve a linear system to calculate the conditional probabilities. When a majority of nodes are assigned to multiple modules, this can take as long as calculating the steady-state distribution of random walkers in the first place. For good performance, we therefore try to test as few combinations of local changes as possible. After testing several different approaches, we have opted for a heuristic method in which we first sort the tuples from best to worst in terms of map-equation change and then determine the number of best tuples that minimizes the map equation. The method works well, because good local changes often are good globally." (Esquivel and Rosvall 2011)

[40] See paragraph 4.1.1

[41] It must be noted that Infomap does not provide specific informations about the percentage of affiliation of a single node to the communities it belongs to. It is an information that is out if interest for the algorithm. Thus, different and arbitrary criteria can be established to study how a single node is allocated over different modules.

[42] Other possible criteria were: the number of submissions, the total amount of received monetary resources, number of agents that every agent had a connection with (through common participations in common projects).

projects in which every agent participated[43], and the most coherent way to handle this aspect seemed to be to assess weights to edges. By doing this, it was possible to focus the attention not only on the total amount of participations that every single agent had, but also on who these participations were with. In this way, the algorithm selects the path of the flow not on the basis of nodes' participations, but on the basis of the number of common participations between nodes. This choice seemed to be the most adequate: it was considered more important to focus on agents' collaborations than to centre the attention on agents' ability to activate themselves.

From the point of view of the algorithm's functioning, to use a weight structure that refers to edges instead of making reference to vertices, does not change any essential aspect. The only thing that changes is that formally the flow is not elaborated as a sequence of nodes, but as a sequence of edges.

#### 4.3.2. Teleportation

The last parameter that has to be discussed is the so-called *teleportation*. Rosvall (Lambiotte and Rosvall, 2012) included in the model the possibility that the flow moves also between couples of nodes that are not connected by any kind of edges. In short, before moving to one of the adjacent nodes, there is the possibility that the flow makes a sort of 'jump' that allows it to reach any node in the network. The probability that the flow moves to a node that is not adjacent to it, instead of continuing its path along connected edges, is called *teleportation probability*. Obviously, if this probability is set to 0, the flow will never make a jump over the network.

First of all, there is an important reason to set the teleportation probability different from zero. This reason lies in the fact that since the aim of this work is to investigate the presence of innovative communities, it is not necessary that these communities originated exclusively from participations in the observed regional public policies. These relationships are crucial element and their most important feature is that they were observable, but a huge number of other kinds of relationships with great importance in the formation of innovation could exist. However connections not related to common participations in considered public polices were not observed, but this does not mean that they did not exist.

Setting the teleportation parameter different from zero makes it possible to simulate connections which, as has been said above, can be regarded as those edges that could not be observed even if there are strong reasons to think they existed. E-mails, phone calls, work meetings, conferences, projects different from the ones observed, standard working relations, commercial exchanges, etc., represent possible kinds of relationships that could have had a role in the flourishing of innovative processes, but that were not observable. To leave room

[43] The use of the number of funded projects as the weight measure, allows to focus the attention on the raw level of participation that every agent had. Other criteria, that for examples could regard financial resources or the number of connected participants (through participations in common projects) would have produced different kind of distortions. On the one hand, the number of connected participants depend on the dimension of the project but this does not have specific importance in this context. On the other hand, looking at the financial resources obtained in the context of considered regional public policies is not reliable because the amount of available resources changed over the waves; moreover, this quantity has more to do with means that were available and, at least from a theoretical point of view, it has a relation with the concept of innovation that seems to be weaker than the number of participations.

to include them in the analysis is important because this introduces a higher degree of realism in the models, given that the Infomap methodology relies completely on a process of simulation.
Thus, the teleportation probability has to be set higher than zero, but it is debatable at which level this probability should be fixed. Rosvall and Bergstrom suggest a value of 0,15. This was determined after the study of several networks[44] and it can be taken as a good default level. However, if more specific hypotheses can be advanced, it is meaningful to adjust the teleportation probability in the most realistic way. Unfortunately, in this work it was not possible to fix a teleportation parameter at a level different from the default value. A more coherent value of this parameter (always with respect to this work) will probably be elaborated only in the future. In fact, a sample of the enterprises that participated in the considered regional public policies were interviewed 6 years after the end of the last wave[45]. They were asked to indicate the number of other projects - other with respect to projects belonging to the regional public policies considered in this work - that in the same period they were participating in. These 'other' projects represent exactly what is described above as unobserved connections and, even if in the mentioned survey there are no questions regarding the identity of the actors they were collaborating with, it could give the opportunity to calculate the ratio between considered network innovation projects and other kinds of projects. If this information were available, it would be possible to reconstruct coherently the probability of teleportation. Unfortunately, at the time of writing this thesis, this survey was still in progress. Since it was not possible to fix the parameter at a level different from the default, in all applied models teleport probability was fixed to 0,15.
To conclude, one final methodological aspect regarding teleportation has to be explained. When teleportation occurs, every point in the network does not have equal probability to be the destination node of the 'jump'. As seen in the previous paragraph, the movement of the flow along the network depends on the weight structure that is provided. It is the same weight structure that influences the movement of the flow along adjacent edges which determines which edge will be the landing point of the teleportation.

## 4.4. Three models of Infomap analysis

After having described those Infomap features that have the most importance in the context of this analysis, I will be introduce the different models and explain their setting. The majority of settings are the same for all developed models. First of all, each Infomap model will be asked to detect overlapping communities and, concerning the teleportation probability, there are no reasons to fix it at values different from the one described above (0,15). About the weight structure, it has already been explained that weights are attributed to edges, emphasizing the number of common participations that couples of agents had.

---

[44] Lambiotte and Rosvall (2012)

[45] "ELT_reti: Effetti di lungo periodo delle politiche a sostegno di reti di innovatori, attuate dalla Regione Toscana nel periodo di programmazione 2000-2006" (2015). Working group made up of Margherita Russo, Annalisa Caloffi, Federica Rossi and Elena Pirani, in collaboration with Sviluppo Toscana.

Finally, what has still remained excluded, is the Markovian order that is going to be used. Since this is the most interesting aspect and since different considerations can be made about it, three different settings are used. The differentiation among the three different models that have been used relies exclusively on the Markovian order.

### 4.4.1. Memoryless flow of information: model MKV1

The first model I proposed (named *MKV1*) is a model in which the Markovian order is set equal to 1. This means that no information about the provenience of the flow is taken into account. When the flow moves through the network, at every step it faces a set of probabilities of moving in different directions that depends on (in order):

- the teleportation probability;
- the position of the flow;
- the weight structure of edges.

As can be seen, no considerations about where the flow comes from are made, so it can be said that the flow at every step moves independently from its past. There are three reasons for the choice to use a memoryless model. First of all, it is interesting to make a comparison between a memoryless model and second order Markov models. Then, second order Markov models could be built not through the observation of real flows, but through the imposition of hypotheses that allowed to reconstruct a 'fake' flow. It follows that even if it is certainly very interesting to apply second order Markov models, it is necessary to be always aware of the fact that its degree of reliability could be affected by construction limits. From this perspective, even if the memoryless model takes into account less information, it is built exclusively on what was really observed. Finally, there is a strong theoretical motivation. Of course, common participations (that are represented by edges in the network) are active during the period over which the project is active, but there is something more. In fact, even after the end of a project, agents that have participated in it can still continue to have relations among them. In this way, simulating the propagation of a memoryless flow over the whole network is like observing the propagation of an information stream in a context where all edges are open. This is the case of a situation in which all agents maintained active their relations with all other agents they collaborated with. It is like observing the network at the end of the policy cycle, assuming that all edges have continued to be active.
To summarize, the characteristics of model MKV1 are:

- weighted network structure;
- allowed possibility for overlapping communities;
- teleportation probability equal to 0,15;
- first order Markov model (memoryless flow).

### 4.4.2. Introducing memory into the analysis: MKV2_AS and MKV2_DS models

The second and the third models that I developed (named *MKV2_AS* and *MKV2_DS*) are second order Markov models. This means that the direction that the simulated flow takes at every step depends also on where it was at the immediately preceding step. It is necessary to remember that this condition, which is of extreme interest, has been applied through some hypotheses about information streams that could have occurred in the real network. No flow was observed but, working on the temporal order of projects and participations, it was possible to determine a 'fake' flow. Of course, the fact that this is just one possible representation must be always borne in mind but, at the same time, it has to be remembered that coherent constraints were respected.

The process of creation of the fake flow was implemented through the creation of trigrams. The existence of a flow among ordered groups of three agents with one of these two characteristics was hypothesized:

- all three agents are involved in the same project (within project flows);
- the first agent and the last agent participated in two different, and sequential projects, while the second agent participated in both of these (among projects flows).

#### *4.4.2.1. Terminated projects*

The idea behind the construction of the fake flow is that agents, through participations in projects, first accumulate knowledge and information, and then they spread it over the network. The crucial aspect that emerged during the process of reconstruction of these fake flows, regarded the consideration of the temporal order of projects[46]. Since a flow is interpreted as a stream coming from the accumulation of information in the context of a single project, the hypothesis was immediately accepted that the flow could move between two projects, the first of which terminated before the terminated of the second. The point is that when a project ends, participants will not have any other information from it and so every thing that was learnt in that context could be transferred into another context. Thus, since it has been decided to simulate the presence of all possible flows among projects, it follows that it is automatic to assume the existence of a flow between a terminated project and any other successive one.

#### *4.4.2.2. Non terminated projects*

Further considerations have to be made for projects that are in progress. The question is: *is it possible that an agent that was participating in two projects at the same time, could have determined a stream of information from one project to the other even if the neither of the two had finished?* Here the point lies in the evaluation of the possibility to transfer information from projects that are not yet terminated. In a context where the goal is to figure out all possible information flows that moved from projects to other projects, of course this situation is possible but, while the condition explained in the previous section (see paragraph 4.4.2.1) is more likely, this one has to be regarded with more doubts. If the general and most important assumption is to give

[46] See paragraph 4.2.3. The sense of these hypothesis is that after having finished a project, participants accumulated knowledge and so they can give rise to a flow of information.

representation to flows that emerged from the accumulation of information gained in projects, it immediately follows that at the end of projects agents have the capability to generate this spread. At the same time, it is not obvious to accept the fact that during the participation in a project an agent is able to give rise to a flow. Is it possible to accumulate sufficient information to give rise to a flow, before the project (which is that permits this accumulation) is finished? Of course it is, but it has to be regarded as a different condition and, compared to the other one, it has to be considered a relaxation of it.

The consideration of the possibility to observe the rise of a flow exclusively from terminated projects, has more to do with an idea of accumulation concerning the knowledge of all the working processes, of all the experiences and of all the phases that constituted the project. On the other hand, admitting the possibility of a spreading of information before the project has terminated, seems to have more to do with a concept of acquaintanceship, rather than accumulation of knowledge. Both must be regarded with interest and both can be crucial in a process of flourishing of innovation. They represent two different situations and as a result the choice was to developed two different models reflecting these two different hypotheses.

#### *4.4.2.3. Weight structure of reconstructed fake flows*

Before presenting the two models, a final consideration has to be made. In these models, a kind of 'homogeneity' condition is assumed: independently of which agent is involved and independently of which project is considered, a single fake flow has always weight equal to 1. From the perspective of the construction of trigrams, this means that each combination of three agents that respects the conditions described above, is calculated as having a weight one. Only if some trigrams repeat themselves, then their weights change, becoming equal to the number of their occurrences.

#### *4.4.2.4. Memory and admitted simultaneity: model MKV2_AS*

In the first of the two models with a second order Markov condition, the propagation of flow moving from projects that still have not terminated is considered. The name *MKV2_AS* stands for: <u>M</u>ar<u>k</u>o<u>v</u> order <u>2</u> with <u>A</u>dmitted <u>S</u>imultaneity[47]. To summarize, the characteristics of this model are:

- weighted network structure determined by trigrams;
- allowed possibility for overlapping communities;
- teleportation probability equal to 0,15;
- second order Markov model;

[47] The term 'simultaneity' refers to the agents' capability (or not) to generate flows from ended projects.

- admitted flows from projects that have not yet finished[48].

#### *4.4.2.5. Memory and denied simultaneity: model MKV2_DS*

In the second of the two models with a second order Markov condition, the propagation of flow moving exclusively from projects that have finished is considered. The name *MKV2_DS* stands for: Markov order 2 with Denied Simultaneity[49]. To summarize, the characteristics of this model are:

- weighted network structure determined by trigrams;
- allowed possibility for overlapping communities;
- teleportation probability equal to 0,15;
- second order Markov model;
- admitted flows only from projects that have already finished[50].

---

48 In the process of construction of the fake flow, it has been made a comparison between all possible couples of projects to determine, with respect to their dates of beginning and of end, which were the cases where it could be observed a flow of information from one to the other. This depend obviously on the hypothesis that is selected. In the case of model MKV2_AS, the condition was:

the beginning of the first project < the end of the second project

Under this condition, every participant could give rise to a flow of information starting from the moment of beginning of the project in which it is involved. Obviously the spread of information could occur only in direction of the later projects.

49 The term 'simultaneity' refers to the agents' capability (or not) to generate flows from ended projects.

50 In the case of model MKV2_DS, the condition was:

the end of the first project < the end of the second project

Under this condition, every participant could give rise to a flow of information starting from the end of the project in which it is involved. Obviously the spread of information could occur only in direction of the later project.

# 5. DCI

## 5.1. Cluster Index: concepts and problematic aspects

Dynamic Cluster Index analysis (DCI) takes its origin from the neurological studies of Giulio Tononi[51]. The research of the Italian neuroscientist and psychiatrist, concerns the investigation of the activity of brain regions, with particular attention to the detection of functional groups of neurons. The idea of Tononi *et al.* (1994, 1996, 1997, 1998) is that there are groups of neurons that show a correlated activity, independently of their position inside the cerebral area.

### 5.1.1. Tononi's studies on functional groups of neurons in the brain area

To suppose that neurons that are situated in brain regions far from each other can have a high level of coordination in their behavior over time, was in contrast with the most important theories about the brain's mode of operation. In the field of neurological activity, two theories have always been opposed: the first, a localizationist theory sustains that the brain is divided into separate areas characterized by specific functions, while the second sustains the presence of a holistic scheme of brain activity. Neither of these formulations were compatible with the hypothesis of the presence of groups of neurons that, regardless of their position, have specific and common functions. To go beyond these theories, Tononi introduced the notion of functional cluster, defining it as "a set of elements that are much more strongly interactive among themselves than with the rest of the system, whether or not the underlying anatomical connectivity is continuous" (Tononi 1997). After taking advantage of some information-theory concepts he conducted a series of studies to detect evidence of the presence of what he supposed to exist: groups of neurons, situated in separate brain regions, that show integrated behaviors over time.

### 5.1.2. Information theory: entropy, mutual information and integration

Tononi's hypothesis about brain functioning concerned the presence of groups of neurons that interact among themselves more intensively than with the rest of the brain. In other words, the point investigated by Tononi *et al.* was to verify the existence of groups characterized by an internal exchange of information (among neurons belonging to the same group) stronger than the exchange of information that the same neurons have with the rest of the system. Looking at coordination and functionality as features densely linked to the presence of information exchange led these neurological studies to borrow concepts from information theory. In particular, what seemed to be extremely pertinent, was making use of the concepts of *integration* and *mutual information* to investigate functional behavioral patterns among neurons. Thanks to these measurements Tononi *et al.* (Tononi *et al.* 1994) introduced

[51] Neuroscientist and psychiatrist. He holds the David P. White Chair in Sleep Medicine, as well as a Distinguished Chair in Consciousness Science, at the University of Wisconsin.

a new concept, the *cluster index*, and found evidence of neurons that could be considered to belong to specific sub-systems, even if they do not belong to the same cerebral area. Now, to better understand Tononi *et al.*'s contributions, it is necessary to introduce some basic information theory concepts upon which the *cluster index* is based. The analysis of the level of coordination of the activity of a group of neurons can be done through the joint probability density function of the neurons belonging to the considered group. Assuming that the activity of this group is described by a stationary multidimensional stochastic process (Shannon and Weaver 1949, Papoulis 1991), the description of this process can be in terms of *entropy* and some other notions related to it.

The Shannon entropy (H) (Shannon and Weaver 1949, Cover and Thomas 2006) of element $i$ can be estimated assuming the frequencies of the states that element $i$ takes as proxies of the probabilities of the states. So, if the states that the *i-th* element takes are denoted by letter $v$, and if frequencies of each state are written as $f_v$, the entropy $H$ of element $i$ is defined as

$$H_i = -\sum_{v=1}^{m} f_v log f_v$$

Moving from the entropy of a single element to the entropy of a subset $X_{jk}$ of elements $k$, means to consider the frequencies of the combined states of elements belonging to $X$. So, if elements that are present in $X_{jk}$ always show the same combination of values, entropy is null, while if $X_{jk}$ is composed of independent elements, its entropy is maximal. The joint entropy of a subset $X_{jk}$ composed of two elements $x_1$ and $x_2$ can be defined as

$$H(x_1, x_2) = -\sum_{x_1}\sum_{x_2} f(x_1, x_2) log f(x_1, x_2)$$

Now that this definition is clear, the two quantities that are crucial in the computation of the *cluster index* can be defined. These quantities are: *mutual information* and *integration*.

*Mutual information* (MI) represents the deviation from the independence of a subset ($X_{jk}$) - where $k$ is the number elements which compose the subset $X_{jk}$ - of the entire system ($X$) from the rest of the system ($X$-$X_{jk}$). It can be thought of as a measurement of the coordination between a subset and its complementary set, and it can be formally defined as

$$MI(X_{jk}; X - X_{jk}) = H(X_{jk}) + H(X - X_{jk}) - H(X)$$

where $H(X_{jk})$, $H(X - X_{jk})$ and $H(X)$ are respectively the entropies of the subset $X_{jk}$, of the complementary set of $X_{jk}$ and of the whole system $X$. MI assumes values equal to 0 in the case where the two subsets are statistically independent. When this happens, the entropy of the whole system matches the sum of the entropies of the two subsets. By contrast, when the two subsets are not independent, the entropy of the system (in which the subsets are considered jointly) is lower than the sum of the entropies of the two subsets.

The last notion that it is necessary to explain in order to introduce the concept of *cluster index*, is *integration* (I). I of a subset $X_{jk}$ can be defined as the deviation from statistical independence among the $k$ elements of subset $X_j$. It can be formally defined as

$$I(X_{jk}) = \sum_{i=1}^{n} H(x_{jik}) - H(X_{jk})$$

and it can be thought of as a measurement of the extent to which the agents of a specific subset are coordinated among themselves.

### 5.1.3. Tononi's cluster index and following developments

Cluster index (CI) can be defined as the ratio of the two quantities introduced above: MI and I. Formally, the CI of *Xjk*, where $X_{jk}$ is a subset of the whole system $X$, is defined as the ratio between I of subset $X_{jk}$ and MI of subset $X_{jk}$. So, CI of $X_{jk}$ can be written:

$$CI(X_{jk}) = I(X_{jk})/MI(X_{jk}; X - X_{jk})$$

Thus CI can be thought of as a measurement of the extent to which elements of a specific subset interact more intensively among themselves than with the rest of the system. As Tononi et al. (1997) explained well "a cluster index near 1 indicates a subset of elements that are as interactive among themselves as with the rest of the system. A cluster index much higher than 1 indicates instead a subset of elements that are strongly interactive among themselves but weakly interacting with the rest of the system, i.e., a set of elements that corresponds to the notion of a functional cluster. Finally, a cluster index smaller than 1 indicates a subset of elements that are less interactive among themselves than with the rest of the system."

The introduction of CI immediately gained a lot of attention because of the results it led to. In particular thanks to some neuroimaging data set collected thought Positron Emission Tomography (PET) and Functional Magnetic Resonance Imaging (fMRI), Tononi *et al.* (Tononi *et al.* 1998) were able to demonstrate the presence of subsets of coordinated groups of neurons. Following these studies, Villani *et al.* (Villani *et al.* 2013a, and Villani *et al.* 2013b) borrowed the concept of CI introducing it in research areas of artificial network models, of catalytic reaction networks and of biological gene regulatory systems, giving an important and recognized[52] contribution to the problem of identifying emergent meso-level structures. Villani *et al.* have investigated the presence of emergent patterns of activity of agents, pursuing two specific aims:

- to test the effectiveness of CI to identify subsystems in artificially designed network (whose nodes were made to follow specific behavioral rules);
- to use the method to study some real complex systems, like chemical or biological systems.

Villani *et al.* first of all developed the concept of CI in the attempt to overcome the problem that CI values depend on the size of the subsystem that is under analysis. Quoting Villani *et al.* (Villani *et al.* 2013b):

[52] Winner of the 'Best Award Paper Award' at European Conference on Artificial Life (ECAL) 2013.

> C(S) scales with the size of the subsystem, so a loosely connected subsystem may have a larger index than a more coherent, smaller one: to compare the indices of the various candidate clusters it is therefore necessary to normalize their cluster indexes, for example by comparing them with those of subsystems having same size, but belonging to a non-clustered homogeneous system (a 'null system'). (Villani *et al.* 2013b)

The scaling problem was overcome comparing the *CI* of the considered subset with the average *CI* of artificial subsets. These artificial subsets all have the same dimension and are all extracted from a homogeneous system[53] in which nodes have a random generated behavior in accordance with the probability of the states they assume in the original system. The definition of a homogeneous system is fundamental to compute a normalization of *CI* and, after that, to calculate its significance[54]. Thus, the normalized cluster index $CI'(X_{jk})$ of a subset $X_{jk}$ belonging to the system $X$, is calculated as

$$CI'(X_{jk}) = \frac{I(X_{jk})}{\langle I_h \rangle} / \frac{M(X_{jk};X-X_{jk})}{\langle M_h \rangle}$$

where $\langle I_h \rangle$ is the average *integration* of subsets (of dimension $k$) belonging to the homogeneous system, and $\langle M_h \rangle$ is the average mutual information of subsets[55] (of dimension $k$) belonging to the homogeneous system. Then, after calculating the normalized $CI'$, its distance ($t_{ci}$) from the homogeneous system is calculated in terms of the standard deviation of the *CI* values that were observed in those subsets (of the same dimension) extracted from the homogeneous system:

$$t_{ci} = \frac{CI'(X_{jk}) - \langle CI'_h \rangle}{\sigma(CI'_h)}$$

where $\langle CI'_h \rangle$ and $\sigma(CI'_h)$ are respectively the mean and the standard deviation of normalized cluster indices of subsets that have the same size of $X_{jk}$ and that belong to the homogeneous system.

### 5.1.4. Comparison among $t_{ci}$ of different subsets

The calculation of the CI significance ($t_{ci}$) in terms of distance from a homogeneous system (distance measured in terms of standard deviation) makes comparable, in terms of $t_{ci}$, every possible subset that belongs to the same set. Thus, regardless of the dimension of the considered subsets, it is possible to compare all subsets belonging to the same set.

[53] The homogeneous system has the same dimension of the original system and every node matches, in term of probability of its states, one node of the original system.

[54] In this context, the word 'significance' does not refer to an inference analysis. Significance must be understood as a measure from the average score, in terms of standard deviation.

[55] The number of subsets that, after being extracted from the homogeneous system, are used to calculate $\langle I_h \rangle$ and $\langle M_h \rangle$, is equal to the number of identified subset (of the same dimension) belonging to the original system.

$$A \subseteq X$$
$$A \subset Y$$
$$X \cap Y = X$$
$$X \cup Y = Y$$
$$X \neq Y$$

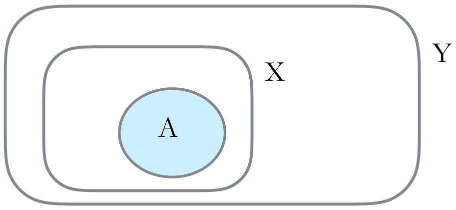



*Figure 5.1. Graphical representation of a particular situation in which tci cannot be compared. Subset A is included both in set X and in set Y. However, since X and Y are different, the tic of A calculated with respect to set X is different from the tci of A calculated with respect of set Y, even if (obviously) A is always equal to itself and even if X is a subset of Y (thus, all its agents are included in Y). On the rights, formal conditions of the example.*

Nevertheless, it remains impossible to make a comparison among groups belonging to different initial sets: the construction of the homogeneous system, a necessary step in the described procedure, depends on the features of the original system. It follows that for every different original set, there is a specific homogeneous system. This makes subsets belonging to different sets totally incomparable.

Finally, it is necessary to fix a specific point about the impossibility to make $t_{ci}$ comparisons. As just explained, it has no sense to make a $t_{ci}$ comparison between two subsets belonging to two different sets, but this is not all. In fact, if A is a subset that belongs simultaneously to set X and to set Y and if set X is perfectly included in set Y, it still makes no sense to make a comparison between $t_{ci}$ of A evaluated in reference to set X, and $t_{ci}$ of A evaluated in reference to set Y. Formally, even under these assumptions $t_{ci}$ of A evaluated in reference to set X, and $t_{ci}$ of A evaluated in reference to set Y, are not comparable. This means, that if the same group of agents is observed in two sets, the latter of which is an extension of the first, the value of $t_{ci}$ calculated with respect to the these two sets still cannot be compared.

To summarize, the comparison of different subsets in terms of $t_{ci}$ can be done exclusively if they all belong to the same set, and only if $t_{ci}$ is calculated with respect to the same set. It is very important to understand the necessity to determine a single set to apply the methodology to. It is explained later[56] as this consideration is crucial in the development of the analysis I proposed in this work.

## 5.2. DCI algorithm: concepts and problematic aspects

The specific contribution of Villani *et al.* lies in particular in the definition of a procedure that allows the ranking of all possible subsets Xjk contained in set X, in terms of $t_{ci}$. As described above the only information necessary regards the behavioral profile[57] over time of

[56] See paragraph 5.5.1.6

[57] For 'behavioural profile over time' is intended a description of the activity of an agent. Obviously this description can be declined in many ways, the simplest of which is describing when an agent is active and when it is not active.

agents involved in the whole set[58]. This procedure, which Villani *et al.* have called *Dynamic Cluster Index* analysis (DCI), involves a series of criteria for the selection of subsets belonging to the whole system.

To find the best subset in terms of $t_{ci}$, every possible subset $X_{jk}$ of system X should be analyzed but, as the number of nodes of the considered system increases, the computation of all possible combinations takes a reasonable time only if the initial set has limited dimensions. This condition can be certainly met in artificially designed networks like those that Villani *et al.* (Villani *et al.* 2013a, Villani *et al.* 2013b, FIlisetti *et al.* 2011) initially used to test the efficacy of the method. By contrast, in biology or in chemistry or in socio-economics it is often of little interest to analyze small systems. For this reason, in order to overcome the problem of the computational duration of an exhaustive $t_{ci}$ calculation, specific steps were introduced in the procedure. Rather than using a random sample analysis for each possible dimension[59], they introduced a heuristic investigation.

Starting with a random sample of 1.000 subsets of the minimum considered dimension[60], DCI algorithm finds the one with the highest score in terms of $t_{ci}$ and then it moves to the analysis of all subsets containing one more agent. In general, to analyze the dimension $k+1$, DCI algorithm considers all subsets of dimension $k+1$ that perfectly include the best subset of dimension $k$. So, moving to the next dimension, all the subsets include the subset detected in the previous step. As described by Villani *et al.*, this heuristic procedure "*has proven to be quite effective, because usually the subsets with highest DCI value are composed of subsets which in turn have a high DCI value, compared to the subsets of the same cardinality*" (Villani *et al.* 2015). Moreover, to ensure a higher level of coverage in the number of evaluated combinations for each dimension, a genetic algorithm was introduced. All these masks (using the language of physicists, '*mask*' is the term used to describe the concept of '*subset*') are subjected to the application of a genetic algorithm that forces crossovers[61] between random couples of these masks, and that produces couples of new masks that are 'sons' of the two to which crossover was applied. This procedure produces 'generations' of masks all with $k+1$ agents

---

[58] "It is important to emphasize that no other information is needed; therefore, the method can be applied even if nothing is known about the structure of the system, e.g., (functional) relations among its variables" Villani et al. 2015

59 "In absence of specific heuristics guiding the sampling, uniform sampling can be chosen. More precisely, a maximum number of samples to be evaluated per subsystem size is defined. In this way, for subsets of small size the error introduced by sampling is rather limited. Indeed, the information provided by RSs of small size is usually more significant than that of large-size because large candidate RSs can be composed on the basis of smaller ones. Nevertheless, a uniform sampling might miss subsets with high DCI value. To overcome this problem, random sampling is complemented with a simple heuristic search" (Villani and Filisetti and Fiorucci and Roli and Serra 2015)

[60] In this work, the minimum dimension was 2. Since the number of the agents considered in the analysis is 352, it follows that the number of possible subsets of dimension 2 is 61.776. The sampling procedure is necessary, otherwise the computational time that it has to be taken for an exhaustive analysis it would be too long.

[61] Supposing that there are 2 masks of dimension k=4, A and B. A is composed by agents (3,5,8,10) and B is composed by agents (2,5,8,11). Through the procedure of crossover is selected a random point where both masks are cut, and then the generated segments are combined among them. If for examples the crossover cuts the masks in two segments, one of length 3 and one of length 1, the results will be mask C=(3,5,8,11) and mask D=(2,5,8,10). In this way masks are merged among them and, after *g* repetitions of the crossover procedure (in this work *g*, that stands for 'generations', was set to 25) it is calculated which is the best of dimension k, obviously in term of $t_{ci}$. Then, the best mask is used to start the same procedure with the dimension k+1.

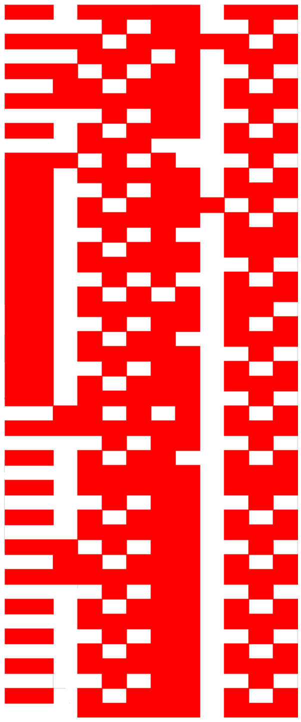
and, after a specific number of generations (in this work, DCI genetic algorithm was set to produce 25 generations for every dimension from 2 to 90), it is evaluated which is the best mask of dimension $k+1$ in terms of $t_{ci}$.
The described procedure continues until the dimension of N-1 is reached, where N is the dimension of the whole system, or until the maximum dimension specified by the researcher is reached. Then, the last thing the algorithm does is to compile a final $t_{ci}$ ranking in terms of all considered combinations of nodes. In addition to all this, since there are random steps[62] that can affect the continuation of the procedure, the whole algorithm can be run several times from the beginning. Each run is repeated with the same principles but it could take different paths, which could lead to different results. As the last run finishes, a complete ranking of all subsets detected during the whole procedure is drawn up. Obviously, the more runs are carried out the more significant are the results: repeating runs increases the probability that the best subset detected through DCI analysis is the best subset in terms of $t_{ci}$ in absolute terms.

### 5.2.1. Potentiality of DCI analysis

After having understood the steps that characterize DCI analysis, it is necessary to investigate it in greater depth it to grasp its functioning and, the most important thing, to understand the potential that this methodology can express. In order to start DCI analysis, the only information that is necessary is the profile of activity of the agents that lies in the system under observation. The behavior of these agents can be easily characterized by a number of *T* variables describing their states, where *T* is the number of instants in time over which the system is observed.
The simplest types of variables that can describe the activity of actors are boolean variables: they are binary variables that assume value 1 if the node is active in the corresponding instant of time, and that assume value 0 if the node is not active in that particular instant. The figure 5.2a shows a representation of the kind of information needed to apply DCI methodology: a dataset containing minimum boolean information is sufficient to run the analysis. On the left, red cells describe the activity status of nodes over time, and on the right results can be observed: in descending order black cells on the same row reveal which are the combinations of agents with more significant values of CI (figure 5.2b).
It is important to focus attention on the fact that without using any kind of information about the topology of the network (figure 5.2a), DCI analysis is able to detect the presence of those subsets in which nodes are effectively organized. Now it is easy to understand the strong potential of this methodology has: it makes possible the investigation of the presence of integrated actors, in respect of their behavior, without knowing the relational structure through which they are linked. Obviously, it should not be forgotten that the example described in figures 5.2a/b/c concerns an artificial network in which nodes have rules that guide their behavior. Arrows and labels ('XOR', 'AND' & 'NOT') that can be observed in figure 5.2c, describe how every node behaves in possible situations that can happen. In the example, there are precise behavioral guidelines and thanks to this, the observable boolean

[62] The random steps that are included regard the initial sample of subsets of minimum dimension and the application of the genetic algorithm.

(a)



(b)


(c)

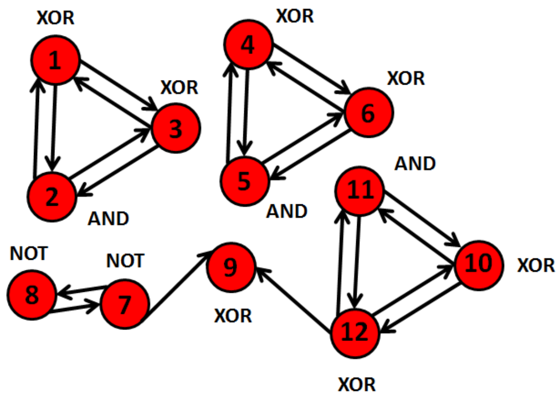



*Figure 5.2a. Graphical representation of the matrix of data. Every column represents one of the 12 agents that belong to the system and the rows represent (in chronological order from the top to the bottom) the 48 instants in which the system was observed. A boolean information was used. Thus, a red cell indicates that the corresponding agent was active in that particular instant in time. A white cell indicates no activity..*

*Figure 5.2b, Results of the application of DCI algorithm. Every column represents an agent (a Node) and they are ordered from the first to the twelfth. The rows represent the masks (subset of agents) detected by the algorithm. Masks are ordered in descending order of $t_{ci}$, with the one with highest $t_{ci}$ on top. If the cell is black, the corresponding agent is included in the corresponding mask.*

*Figure 5.2c. The artificial network that was used to generate the initial matrix of data (Figure Xa). Nodes have rules that guide their behavior. Arrows and labels ('XOR', 'AND' & 'NOT') describe how every node behaves in possible situations that can happen. A random simulation was performed and boolean status were observed and then were used to test the DCI analysis.*

*DCI algorithm, without using any kind of information about the typology of the network, was able to detect the groups that were present in the artificial model. The best mask detected includes node N8, N9, N10, N11 and N12, nodes that are truly connected in artificial network. The second best mask is an extension of the first, since there is the addition of N7. The third mask is made up by N4, N5 and N6 and the fourth is made up by N1, N2 and N3. Also these masks detected groups that were observable in the artificial network.*

pattern created out of a computer simulation shows clear evidences of integration among agents: even the human eye[63] can see integration looking at the last five columns of figure 5.2a.

Finally, it is important to underline a particular feature of DCI analysis. In saying that its main characteristic is to detect groups of agents that show an emergent behavioral pattern, it does not mean that those agents have to behave in the same simultaneous way over time. Of course, this is a possibility but the concept of 'pattern' depicts a situation in which a structure can be observed, a repetition of combinations of individual behaviors. In figure 5.3

[63] Necessarily has to be said that in this example the order of columns facilitates the interpretation of the image, but the point that has to emerge is that the system under analysis was thought and created as a very simple design and a very clear structure.

the behavioral pattern of the best mask described in column 8, 9, 10, 11 and 12 in figure figure 5.2a is highlighted. The columns, each of which represents the activity profile of a single agent, reveals clearly a unique pattern where agents do not have the same behavior in a simultaneous way. Looking at the columns, it is very apparent that there is a high level of integration among those agents, but what should be noticed is the presence of three different combinations:

- agent 8, agent 10, agent 11 and agent 12 are simultaneously active while agent 9 is not active;
- agent 8 and 11 are simultaneously active while agent 9, agent 10 and agent 12 are not active;
- agent 8, agent 10 and agent 12 simultaneously active while agent 9 and agent 11 are not active.

Now it is clear that to determine a behavioral pattern there is no need to have agents that over time behave in the same manner (all together simultaneously active or all together simultaneously not active). It may be possible that agents behave in the same way simultaneously, but this is not necessary. In the end, what really matters is the repetition of particular combinations, a fact that allows an integration - that in the end can be seen as the repetition of specific combinations of agents' activity - in the activity schemes of agents belonging to the same subsystem to emerge. It is very important to underline this capability of DCI because the integration of agents' behaviors must not be confused with the idea of agents doing the same thing at the same time. Even more, this feature demonstrates the high potential of DCI methodology.

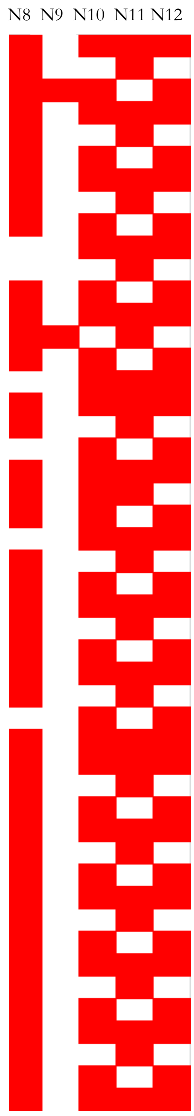



*Figure 5.3. Particular of the matrix of data represented in Figure Xa. Here only the last five columns were represented. These column represent the activity status of nodes N8, N9, N10, N11 and N12 in a boolean form (red cells represent the presence of activity, white cells means no activiy). The rows represent (in chronological order from the top to the bottom) the 48 instants in which the system was observed. This is the best mask which the DCI algorithm detected in that analysis. It can be noticed the integration of behaviors that occurs among these agents. The three combinations of behaviors that occur the most are:*

- *agent 8, agent 10, agent 11 and agent 12 are simultaneously active while agent 9 is not active;*
- *agent 8 and 11 are simultaneously active while agent 9, agent 10 and agent 12 are not active;*
- *agent 8, agent 10 and agent 12 simultaneously active while agent 9 and agent 11 are not active.*

*Having so much stability in terms of combinations of agents' activities is what is intended by an integrated behavioral pattern.*

### 5.2.2. Introducing DCI in a socio-economic context of analysis

Without any doubt, in the field of biology, chemistry and physics, DCI methodology has shown itself to be an incredibly fascinating tool of analysis. Considering new possible fields of application, it would seem to be able to maintain the same strong potential in terms of what it could reveal. One of the major objectives of this work is to introduce this methodology in a field of application completely different from the one in which it was created. Without any doubt trying to introduce DCI analysis into social and economic sciences represented a juicy challenge under many aspects: from the theoretical elements to its implementation.

Before discussing how it has been implemented, a particular theoretical aspect has to be considered. This point regards the nature of the system and of the agents involved in the analysis. In fact, moving to a socio-economic context necessarily involves three different considerations:

- first of all, it is more difficult to isolate the system under analysis, and since agents can be influenced by external elements, a high degree of openness can compromise the analysis enhancing the difficulty to investigate the dynamic of those who are in the focus of the analysis;
- then, while in the context of artificial and chemical systems initial conditions can be easily established, making it relatively easy to understand the relation between context and agents, this is much harder to do in socio-economic systems. Here, initial conditions are difficult to control and, because their unavoidable influence on observed behaviors, they can lead to misleading interpretations. Only a deep comprehension of what has happened can ensure a clear evaluation of the present, but this is not frequently the case;
- finally, there is also an ontological element that in this kind of field makes the analysis so delicate. Considering socio-economic actors means to deal with agents that may have behavioral schemes, but that do not necessarily follow precise rules of interaction.

After these considerations, it is easy to understand that introducing DCI analysis in the socio-economic field has immediately and necessarily to do with the comprehension of delicate aspects. Being aware of the differences between the study of human behavior in social organizations and the study of molecular reactions in the evolution of life, is a necessary condition in the process of investigating the presence of innovative communities with this new methodology, as I proposed here.

At the same time, the aspects listed above and that have to be borne in mind, should not prevent the application of DCI methodology to a field different from the one it was created for: even if differences have been highlighted, common elements are present and they justify this attempt of investigation. In particular, the points that give sense to the use of DCI technique in a such different context, are:

- the presence of agents that have recognizable behavior over time;
- the intent to detect the presence of organizational structures ('*subsystems*' in physics, and '*communities*' in social network analysis);

- the conviction that an essential feature of organizations lies in the integration of agents' behavior, an element of cohesion that is adequate to make suppositions about the presence of shared and common functions.

While the first assertion is a consideration on the available data structure, and the second assertion is a simple declaration of the presence of a strong similarity between one of the objective of this work and the objective that inspired the development of DCI technique, the last point highlights one of the elements that I intended to investigate through this work. The literature that was proposed in paragraph 1.2 concerns a specific vision of organizations - a concept that can be considered extremely close to that of an economic community - that stresses the necessary presence of a common function. Agents belong to an organization if, in addition to having a relational structure and to having the ability to manage processes, they share one or more specific targets that guide the activity of the whole group. The element that necessarily has to emerge is the presence of a purpose, an objective thanks to which the community finds its own specific sense. Without investigating the specific and detailed purposes of communities, my attempt is to evaluate the presence of a common objective which, if it exists, is likely to imply the presence of emergent behavioral patterns.

### 5.2.3. DCI and unobserved relations among agents

There is another consideration about DCI analysis that deserves attention. As described above, DCI analysis allows researchers to investigate the presence of communities of agents without any kind of information that regards the topology of the network. This particular feature immediately captivated my attention, not only for the potential that it expresses, but especially for the perfect accord with the specific aim of this work.
I have described that the aim of this thesis is to explore ways to investigate the presence of innovation communities and clarified that the available data regard participations in regional public policies. Of course, what cannot be denied is the coherence between the object under investigation and the informational basis upon which the analysis has been set up. However, at the same time, it is not intellectually correct to think that innovation processes could have arisen only through the public policies under investigation. The network that has been elaborated and then analyzed with different methodologies, is exclusively based on participations in these policies. Thus, the question is: is it not possible that innovation processes could have bloomed not only through common participations in funded projects?
The answer is affirmative: it is possible. Of course, it is important to point out again that collected the data represent an incredible resource, but it also necessary to comprehend that with respect to the aim of this work, an analysis that focuses its attention only on relations generated by funded projects seems to lack something. It is likely that agents did not have relations among themselves exclusively on the basis of common participations in funded projects and, consequently, it is not unreasonable to think that:

- unobserved relations could have been important in agents' decisions about participations in regional public policies;

- unobserved relations could have had a role in the process of the flourishing of innovation.

E-mails, phone calls, work meetings, conferences, projects different from the ones observed, standard working relations, trading operations, etc.: these are some possible examples of relations that could have had a role in generating innovative communities. And from the point of view of this work, it would have been incredibly interesting to consider and to study these relations, but they remained unobserved.
To overcome this lack, the intuition was to apply DCI methodology to this specific case of study to nullify the presence of network information in the analysis. Since it has been clarified that there could have been different kinds of unobserved relations, DCI is seen as an opportunity to investigate the whole set of relations that were active during the period under observation. All relations - and not only relations generated by common participations in funded projects - could have had a role in determining behavioral patterns of activity in the context of these regional public policies. Thus, DCI represents an incredible technique to deepen the analysis of what happened, moving beyond the fundamental but limited consideration of common participations in projects. From this perspective the importance of DCI in the context of this work seems to have all the requirements to lead to a wider range of considerations and to new wider opportunities of analysis.

## 5.3. Managing data in order to apply DCI

In order to apply DCI analysis to this specific case of study, an extensive appraisal of the available data structure was carried out. In this section, the domains of these interventions are discussed step by step.

### 5.3.1. Typologies of variables

As illustrated above, what DCI analysis requires, is a series of variables describing agents' activity over time. Since the object under investigation is a series of public policies, the choice that was made was to characterize agents' behavior in terms of participation in projects involved in the considered public policies. An agent was defined to be active during those instants in which it was participating in at least one funded project, and since projects' starting and finishing dates were available, a complete profile of activity for every agent involved was defined. Obviously, by contrast agents were considered to be inactive during those moments in which they were not participating to any funded projects[64].
The first question it was necessary to address, concerns how to define agents' activity. Different solutions were proposed:

---

[64] It is important to stress the fact that, at the beginning of this stage of the work, the decision was to completely exclude information about not-funded projects. It is important to remember that these projects were not evaluated as non admissible. They passed the first phase of evaluation in the process of admission to funds, but at the end they were not funded because other projects had a better evaluation. To conclude, non funded projects are not projects that were rejected.

(i) a series of boolean variables that assume value 1 if the agent is active and value 0 if it is not, surely the simplest way to describe an activity status (variables *bool_**)[65];
(ii) a series of discrete variables, one for every observed moment in time, whose value is equal to the number of funded projects which the agent is participating in at that moment (variables *lev_act_**);
(iii) a series of discrete variables, one for every observed moment in time, whose value is equal to the sum of participants that every agent was supposed to be in contact with, through its participation to funded projects in the corresponding moment of time (variables *connAg_**);
(iv) a series of categorical variables, one for every observed moment in time except the first one, describing the relation between the number of funded projects the agent is participating in at the corresponding moment, and the number of funded projects the agent was participating in at the previous moment( variables *dloa_**).

Considering about the selection of information that could reveal agents' integration in terms of behavior, the second and the third options were abandoned, since they were describing the activity status of agents including also information about the number of projects or about the number of participants in projects. The evaluation was made upon the fact that:

- the number of participants in projects that agents were participating in, could not contain information about agents' integration;
- the number of projects that agents were participating in was not convincing because of the fact that it is an expression of a level of activity that has more to do with the single agent's attitude than with the possible presence of integration among agents.

Since the objective of DCI analysis is to capture integration among agents' behaviors, it seemed to be more adequate to use (i) the boolean variables and (ii) the variables expressing the variation of the level of activity. The rationale of this choice lies in the fact that both these variables express in a more evident way the information upon possible agents' integration. On the one hand, looking at boolean information simply puts in evidence a scheme of activation, and so the idea is to check if there are groups of agents that activate themselves with particular patterns. On the other hand, performing the analysis with variables expressing the variation of activity levels seems to be a good solution to understand if there had been some coordination in how agents increased or diminished the activity. This information was also included in the series of variables that were not selected for the analysis but, along with this information, there was also other information concerning elements (like the number of projects or the number of participants) that seemed to have little to do with the integration of agents' behaviors.

[65] The star * indicates series of variables with the same prefix

### 5.3.2. Time dimension

So far, the time dimension has generally been referred to using the term 'moment'. To explain how the time dimension was managed, it is necessary to add one more consideration about DCI analysis. As has been shown above, the integration of agents is pursued by observing their activity over time and it follows that the longer it is possible to observe agents, the more accurate the analysis will be. This is true, but only in part.

Because of the fact that DCI is calculated through a comparison between entropy related measurements of the whole set and entropy based measurements of subsets, when the whole set remains equal to itself over time no useful information is added to the analysis. The permanence of the whole set in a particular state does not affect any subset's ratio between the exchange of information with the rest of the system and the internal integration. That is the reason why the crucial element that has to be considered in the construction of time data, is that the highest number of instants in time is required, but under the condition that there is a variation in the states that the whole set assumes.

This fact immediately led to the consideration of the projects' starting and ending date as points of reference in time dimension. Because of the fact that it was possible to reconstruct the agents' time profiles only on the basis of their participation in funded projects, it immediately followed that the only moments in which it was possible to observe a variation of the whole system, were those moments in which at least one project started or in which at least one project ended. Working on all available dates of starting and of ending of funded projects, it was possible to define 52 instants over time, where an instant corresponds to a daily date. Looking at DCI analysis tests[66], and bearing in mind that the system under observation is composed of 1.121 different agents (the number of agents that have participated at least in one funded project), inevitably the number of defined instants seems to be too few. In fact, Villani *et al.* (Villani *et al.* 2013a, Villani *et al.* 2013b) tested DCI analysis on a dataset determined by 12 agents observed in 30 different instants over time. Even if there are not precise indications regarding the proportion between the number of agents and the number of instants in time, it is necessary to observe the system over a number of instants that allows the investigation of the presence of behavioral patterns. Obviously, if the system is determined by a small number of agents, a large number of instants would not be necessary. Unfortunately, this is not the case of this work. The presence of 1.121 agents involved should have been balanced by observing them repeatedly, but this was not possible.

To overcome this problem, before considering forcing a reduction in the number of agents considered to deal with the problem of the proportion between agents and instants, it was first attempted to use information coming from non funded projects to increase the number of instants. In fact, among 1.121 agents that participated in least to one funded project, many were involved in projects that after a process of evaluation were not admitted for financing. Even if it was decided to focus the attention of the whole work exclusively on funded projects, in this case it seemed appropriate to use information about not funded projects to increase the number of observable instants over time. The solution that was adopted was to factitiously consider a positive activity status of the duration of one day for every project that was not funded (and consequently for all those agents who participated in

[66] Filisetti *et al.* 2011

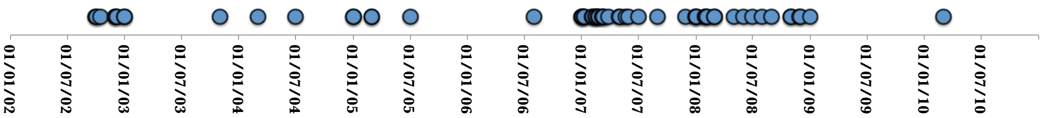



*Figure 5.4 (above). Timeline from 01 January 2002 to 01 July 2010. Over this period the projects were developed. Every point dot is one of the dates of beginning or of the end of at least one projects. The dots are 59 and they correspond to the 59 instants over which the system was observed. It is possible to observe how the dots are not uniformly distributed over the timeline.*

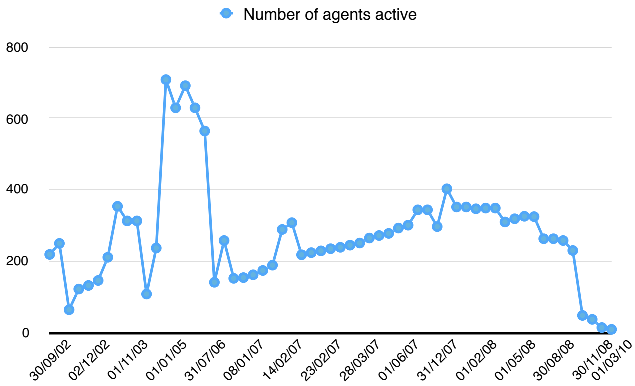



*Chart 5.5 (above). Number of agents active (over the 1.121 that participated in the public policies), in the different 59 instants detected from the observation of the dates of beginning and of end of the projects. The 59 instants are represented in sequential order, with no consideration for distances in time.*

not funded projects). Perfectly aware of the fact that it is an arbitrary decision, this solution does not seem to be meaningless, since the creation of a project is a process that deals with agents' relations and exchanges of information. With the purpose of adding new dates, the choice was to assign to the day before the beginning of the wave to each non funded project as its only day of activity[67]. Because of the fact that out of total of 9 waves[68], there two couples of waves that started on the same day[69], it was possible to add 7 new instants in time, allowing a final extension of the time dimension to 59 daily dates[70].

[67] No funded projects have no specific time information. The only available information regards the earliest date projects could start. This date, that can be considered as the date of beginning of the wave, seemed to be the more adequate reference to fix a positive activity status of non funded projects.

[68] 9 different policy waves are (year and programme): 2002_171, 2002_172, 2002_ITT, 2004_171, 2004_171A, 2005_171, 2006_VIN, 2007_171, 2008_171.

[69] Wave 2002.171 and wave 2002.172 have the same starting date. The same situation for wave 2004.171 and wave 2004.171A

[70] List of final available dates:
30.09.2002, 01.10.2002, 14.10.2002, 01.12.2002, 02.12.2002, 05.12.2002, 31.12.2002, 01.01.2003, 01.11.2003, 29.02.2004, 30.06.2004, 31.12.2004, 01.01.2005, 28.02.2005, 01.03.2005, 01.07.2005, 31.07.2006, 31.12.2006, 01.01.2007, 02.01.2007, 08.01.2007, 10.01.2007, 01.02.2007, 05.02.2007, 14.02.2007, 15.02.2007, 16.02.2007, 21.02.2007, 23.02.2007, 08.03.2007, 09.03.2007, 13.03.2007, 28.03.2007, 01.05.2007, 02.05.2007, 21.05.2007, 01.06.2007, 01.07.2007, 01.09.2007, 30.11.2007, 31.12.2007, 01.01.2008, 02.01.2008, 31.01.2008, 01.02.2008, 04.02.2008, 29.02.2008, 01.03.2008, 01.05.2008, 01.06.2008, 30.06.2008, 31.07.2008, 30.08.2008, 30.10.2008, 31.10.2008, 27.11.2008, 30.11.2008, 31.12.2008, 01.03.2010

| Number of funded projects in which the agent participated | Number of agents | % of reduction in the number of agents, in respect of the immediately upper level |
|---|---|---|
| ≥ 1 | 1.121 | |
| ≥ 2 | 352 | *68,6%* |
| ≥ 3 | 180 | *48,9%* |
| ≥ 4 | 105 | *41,7%* |
| ≥ 5 | 68 | *35,2%* |
| ≥ 6 | 50 | *26,5%* |
| ≥ 7 | 33 | *34,0%* |
| ≥ 8 | 24 | *27,3%* |
| ≥ 9 | 13 | *45,8%* |
| ≥ 10 | 10 | *23,1%* |

*Table 5.6. Number of agents (second column) that had at least the corresponding number of participations listed in the first column. In the third column, the ratio between the number of agents that moving level by level are excluded and the total number of agents that were present before the exclusion occurred.*

A final consideration about the time dimension. DCI analysis does not consider the duration of states. This implies that in the process of selection of daily dates as time references, no attention has given to how far an instant is from the next one in terms of time duration. If the 59 instants are represented over a timeline, their distribution is not uniform (figure 5.4).

### 5.3.3. Reduction in the number of agents considered

As already introduced in the preceding section, the application of DCI analysis requires the observation of agents in time as long as possible. The management of data allowed the identification of 59 different instants in time and no more addition could be made since the DCI needs to observe changes in the whole system[71]. This fact, for example, makes completely futile the attempt to consider every single day from the beginning of the first wave as possible instants in which to examine agents' behavior.
The presence of 1.121 agents[72] that participated in at least one funded project, immediately brought to light the problem of a disproportion between the number of agents and the number of instants. Former tests[73] on DCI never faced the problem of limited number of observations over time, but from the beginning of this work it seemed evident that passing from a ratio of 30/12 (number of instants in time and number of nodes) in the artificially designed network to a ratio of 1.121/59 in the specific case study of this work, would caused some kind of distortion in the analysis[74]. After a briefing with Marco Villani, Roberto

---

[71] See paragraph 5.1.4

[72] Moreover, another important problem in trying to apply DCI analysis to a group of more than on thousand agents, regards the computational effort (in terms of time) that it implies.

[73] Filisetti *et al.* 2011

[74] It always must be kept in mind that the algorithm starts with a random sampling of 1.000 subsets of minimum dimension. If the minimum dimension is set to 2, and the number of considered agents is 1.121, it follows that the number of possible subsets is 627.760. Trying to reduce the number of considered agents is also necessary to make the whole process of analysis more reliable in respect of the covering of initial exploration of subsets.

| Number of Agents | Agents' Typology | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Number of funded projects in which the agent participated | GI | TA | CC | SC | LGI | E | R | BS | KIBS | **TOTAL** |
| ≥ 1 | 39 | 39 | 11 | 6 | 77 | 607 | 62 | 17 | 263 | **1.121** |
| ≥ 2 | 13 | 10 | 10 | 2 | 42 | 119 | 28 | 7 | 121 | **352** |
| ≥ 3 | 9 | 1 | 9 | 2 | 20 | 31 | 21 | 6 | 81 | **180** |
| ≥ 4 | 1 | | 8 | 1 | 9 | 13 | 16 | 1 | 56 | **105** |
| ≥ 5 | | | 5 | | 8 | 6 | 12 | | 37 | **68** |
| ≥ 6 | | | 5 | | 7 | 4 | 6 | | 28 | **50** |
| ≥ 7 | | | 2 | | 6 | 2 | 5 | | 18 | **33** |
| ≥ 8 | | | 1 | | 4 | 2 | 5 | | 12 | **24** |
| ≥ 9 | | | | | 2 | 1 | 4 | | 6 | **13** |
| ≥ 10 | | | | | | 1 | 3 | | 6 | **10** |

*Table 5.7. Number of agents (distinguished by their typology) that had at least the corresponding number of participations listed in the first column. The variable 'Typology' is made up by ten different categories: Government Institutions (GI), Trade Associations (TA), Chambers of Commerce (CC), Service Centers (SC), Local Government Institutions (LGI), Enterprises (E), Universities and Public Research Centers (R), Private Research Enterprises (PRC), Business Services Enterprises (BS), Knowledge-Intensitive Business Services (KIBS).*

Serra, Andrea Roli and Alessandro Filisetti[75], the suggestion was to reduce the number of agents considered. Moving in this direction immediately led to the problem of selecting a criterion to which making the decision about the exclusion of some agents from the analysis. Possible criteria considered were:

- number of funded projects which agents participated in;
- number of project submissions (without distinguishing between funded projects and non funded projects);
- number of agents with which every agent had a connection through common participations in common projects;
- amount of monetary resources received in the context of considered public policies.

The second, the third and the fourth criteria were considered less appropriate than the first. Trying to reduce the number of agents on the basis of the number of submissions could conceal the real ability to have participated in projects selected to be funded. The third criterion emphasizes the participation in projects with a large number of participants and this could conceal information about the specific activity of the considered agent. Finally, the

[75] Marco Villani for is a member of the Department of Physics, Computer Science and Mathematics of the same university.
Roberto Serra performed research activities in the industrial groups Eni and Montedison, where he served as director of the Environmental Research Centre until 2003. In 2004 he moved to the Modena and Reggio Emilia University where he is full professor of Complex Systems at the Department of Physics, Informatics and Mathematics.
Andrea Roli works at the University of Bologna at the Department of Informatics as researcher scientist Alessandro Filisetti is postdoc at the European Centre for Living Technology in Venice.

last criterion described focuses on the ability to raise funds, but this it is not the priority in the context of this analysis.
In the end, evaluating agents on the basis of the number of times they activated (having defined the concept of 'activation' as the participation to funded projects) seemed to be the most coherent criterion with respect to how it was possible to describe agents' behaviors over time. To decide to exclude agents on the basis of the number of projects they participated in, can be seen as the most coherent way to provide the most interesting input for DCI analysis. In fact, it is necessary to remember that DCI analysis focuses its attention on the changing status of agents and, having defined agents' status on the basis of their participation in projects (see paragraph 5.3.1), inevitably the first listed criterion has to be selected.
After having discussed possible criteria to reduce the number of agents and after having selected one of them, it was necessary to evaluate where to stop excluding them. Table 5.6 illustrates the progressive reduction in the number of agents with respect to the progressive increase in the minimum number of funded projects they participated in. As can be observed, passing from the initial set of agents with at least one participation in funded projects (1.121 agents) to the subset of agents with at least two participations in funded projects (352 agents), immediately permits an important reduction in the size of the initial group. Since the necessity was to cut significantly the dimension of the initial set, more intense cut were also explored. In order to do this, it was important to understand which agents were going to be excluded, especially in terms of their institutional typology. How progressive subsets change in terms of their institutional typology, is described in table 5.7.
Trying to restrict even further the subset of agents with at least 2 participations in funded projects, would have dramatically reduced the number of enterprises (E) which from 119 units would have reduced to 31 units. In addition, also the number of KIBS would have strongly reduced: from 121 to 81. The loss of 40 agents of this typology would have caused an impoverishment of the set under a very important aspect since KIBS have a central role in innovation processes because of their very nature.
These were the elements that prevented applying a deeper cut in the considered system. In fact, even if the communities under study in this work do not have to have a specific theoretical composition in terms of the typology of the agents belonging to them, regional economic frameworks are above all characterized and determined by private enterprises. Even if many of these enterprises did not have a high level of participation in considered public policies, it must be borne in mind that they characterize the regional economy and they participate actively in its process of evolution. Therefore, it was evaluated as acceptable (especially because of the necessity to reduce the initial set) to exclude all those agents that had only one participation in funded projects, but it was decided not to go further because it would have caused the loss of a group of agents whose importance is determinant from an economic perspective.
To conclude, the group of agents considered in all applications of DCI analysis, and also in the application of other methodologies, is the group of those 352 agents that participated in at least two funded projects.

## 5.4. DCI algorithm's output: cluster analysis of masks

An element that has to be introduced is DCI outputs are articulated. As was shown above, the DCI algorithm returns a ranking of groups ordered by CI significance. As long as the number of agents involved in the analysis is small, the output is quite clear and simple to analyze. In the example that was described above (figure 5.2b) the five best combinations of agents were listed and it was easy to determine different groups. However, observing the image carefully, it can be seen that the first and the second masks[76], as well as the third and the fifth masks, differ from each other for the presence/absence of just one agent. In fact, even if DCI algorithm returns a ranking of masks that are different from each other (there is no mask equal to another one), each couple of masks can differ just by one element. This fact may create some problems in interpreting the results of the algorithm, since it is possible that the first N masks differ from the best one just in one element. Very similar masks could head the ranking, making it more difficult to interpret. In addition, the DCI algorithm returns a ranking of 75.000 masks: a tremendous amount of output that has to be managed in order to collapse it to a list that can be interpreted more easily by the researcher.
To do that, I proposed the introduction of a hierarchical agglomerative cluster analysis of the masks, combined with a final selection (for every cluster) of the mask that has the highest $t_{ci}$ - obviously among those that are in that cluster - as the representative mask of the considered cluster itself. The first stage of this procedure regards the creation of a progressive agglomeration of all the masks with respect to a criterion of similarity. A

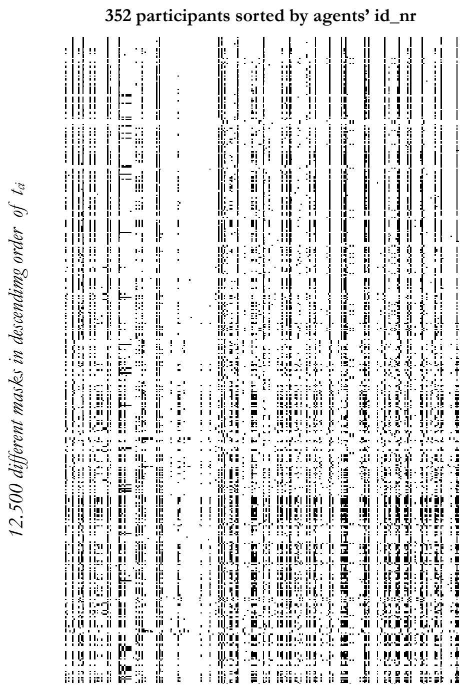



*Figure 5.8. Example of DCI algorithm's output. 12.500 best masks detected. Descending order of $t_{ci}$. The best mask on top. Every column represents an agent and every row represents a different subset. A black cell indicates that the corresponding agent belongs to the corresponding mask. All these 12.500 masks are different, so there is not a pair of identical rows. Every mask can differ from the other just for the presence/absence of one agent.*

[76] '*Mask*' is the term that refers to groups of agents detected by DCI algorithm

similarity matrix is computed calculating the simple matching similarity coefficient[77] for every possible couple of masks. Thanks to this similarity matrix, it is possible to create a dendrogram in which at every step the two most similar clusters merge. Then, among all possible criterions of cluster agglomeration, the complete linkage criterion was used because it merges clusters on the basis of the greatest distance between them[78]. Taking the couple of furthest observations as the representative observations of each considered clusters, guarantees that all other possible couples of observations are necessarily less dissimilar[79].
Then, after the dendrogram is ready, the cut is made evaluating not only the height of branches, but also looking at the composition of the groups, trying to capture the highest number of diverse masks without permitting too high an increase of them. Finally, when the

---

[77] For binary data, like the information about the presence of agents in one community, one possible coefficient of similarity is calculated as the percentage of the number of elements that two observations have in common, out of the total elements that are present in the system. Indicating letters *a, b, c, d* as the four possible combinations of two binary variables, and indicating with capital letter the frequency that the corresponding situation occurs between variables of *i-th* observation and variables of *j-th* observation, it is possible to describe simple matching coefficient as

$$\text{Simple matching} = A + D \,/\, (A + B + C + D)\text{ , where:}$$
$$a=(1,1) \qquad b=(1,0) \qquad c=(0,1) \qquad d=(0,0)$$

Among all binary coefficient, it was selected simple matching coefficient because of two facts:

- denominator is defined as the total amount of variables considered, without regarding if they assume value 0 or value 1. This implies that denominator is always constant among all possible combinations of observations, and this avoids the fact that two small communities have an elevated coefficient just because they are small. This situation could happen using for example Jaccard coefficient, that is defined as
$$\text{Jaccard} = A \,/\, (A + B + C)$$

- it is considered as an element of similarity also agents' contemporary absence from the *i-th* observation and the *j-th* observation. In the case of this work, it means that if an agent does not belong to both *i-th* community and *j-th* community, this situation is evaluated as an element of similarity. Other coefficients, like for example Russell coefficient, consider the presence of similarity just as the contemporary presence of situation *a*. In the specific case of this work, assuming this criterion could have caused some problems like those described below:

| | | | |
|---|---|---|---|
| community A | 1 1 0 0 0 0 1 0 0 1 | community B | 1 1 0 0 1 1 1 0 1 0 |
| community B | 1 1 0 0 1 1 1 0 1 0 | community C | 1 1 1 1 0 0 1 1 0 1 |

| | | | |
|---|---|---|---|
| Russell coeff. (AB) | = 3/10 | Russell coeff. (BC) | = 3/10 |
| Simple matching coeff. (AB) | = 6/10 | Simple matching coeff. (BC) | = 3/10 |

where Russell coefficient = A / (A + B + C + D)

Even if both couples AB and BC have 3 variables that simultaneously assume value 1, between community B and community C all other variables assume different values. In the context of this work it means that communities with just a common base of agents would have been considered as similar as communities with the same common base and a simultaneously absence of others agents.

[78] The complete linkage criterion of agglomeration establishes the similarity of two clusters as the similarity of their most dissimilar members. In other words, those two clusters that have the maximum similarity in their most dissimilar members are merged.

[79] This results in a preference for compact clusters with small diameters over long, straggly clusters, but also causes sensitivity to outliers. A single observation (that is a community in our case) far from the center can increase diameters of candidate merge clusters dramatically, and completely change the final clustering. This is a risk that should not emerge because of the fact that DCI algorithm returns a list of masks that can have minimum variation among them. This makes difficult to observe a single combination of agents very different from all other mask that are present in the final ranking.

clusters are defined, I proposed to assume as the representative mask for every identified cluster, the mask that has the highest score in terms of $t_{ci}$. The application of this criterion means that, among a group of similar masks, the mask that shows the best integration in its agents' behaviors is considered.
This whole proposed procedure follows a clear and distinct principle: the intent to investigate diversity among masks. All masks that are put in the final cluster analysis maintain significance in terms of their level of behavioral integration and so, without being too interested in the best masks of the whole ranking list, I thought it was more interesting to research the subsets of agents that have the highest degree of diversity among them. This direction is due to the nature of the object that is under analysis. Since a socio-economic system is being considered and since it is not known at what phase of evolution its communities are (if they exist), a relatively low $t_{ci}$ does not necessarily indicate that this is the best integration involved agents can achieve together[80]. It is possible that a low significance of CI is due to a phase of decline or to a phase of growth. These communities are not necessarily communities that cannot work with as much functionality as the best identified communities do. They can represent new groups that are only at the beginning of their path, or they can represent groups that after having had a strong integration are now facing a decrease in terms of integrated activity. All these considerations fostered the assumption of a criterion of diversity. The most adequate criterion of interpretation of DCI output, with respect of the case under study, seemed to be to find masks that were diverse rather than performant.
Finally, it is necessary to highlight the fact that, even if moving in the direction of finding masks different from each other, the best mask (among all) in terms of $t_{ci}$ is always considered. Despite privileging a criterion of diversity, a criterion of functionality of masks has also been respected.

## 5.5. Three models for DCI analysis

Every aspect of DCI analysis has been described in the previous sections. At the time this work was starting, DCI analysis was considered a new methodology and it had never been applied to a socio-economic context of research. Since then, many steps have been taken to implement the algorithm, many evaluations have been developed and many choices have been made, in order to understand how the model could better fit a new field of application. From the perspective of this work, it is important to describe carefully the first attempt to apply DCI analysis to this case study. It is important to do it because this first exploration contributed dramatically to the set up of a more definite and coherent analysis. Thus, before describing the two models that have been used in the definitive application of DCI methodology, it is necessary to introduce the first process of analysis and to explain every detail of it.

[80] See paragraph 5.2.2

### 5.5.1. Explorative process of DCI analysis: 'BOOL_1'

The first process of analysis with DCI methodology was characterized by the use of boolean variables[81] to describe agents' activity over time. For this reason, this process will be called 'BOOL_1'. As for all other DCI applications, in BOOL_1 are considered those 352 agents that had at least two participations in funded projects, and these agents are observed in 59 different instants over time. At the time this first process started, DCI implementation had just been developed and it was not set to make automatic repetitions of launches the algorithm. As explained in paragraph 5.2, since the DCI algorithm is characterized by an initial random selection of subsets of minimum determined dimension, it is convenient to repeat the algorithm several times. The fact that the repetition of runs was not automatic has not to be underestimated in particular with respect to some problematic aspects regarding the launching of computational calculations[82]. After the first run of DCI analysis, it became clear that a single run of the algorithm took around five days, and only after following implementations by Alessandro Filisetti the computational time was reduced. Moreover, at that time there was no idea of making a unique final ranking out of several analyses on the same system: every single run of the algorithm produced its own ranking list of masks and it was not merged with other rankings.

These conditions need to be remembered because they had an impact on the process of exploration of DCI. In fact, only after the first analysis was it possible to elaborate questions and formulate needs that it had not been possible to foresee before starting. The comprehension of the initial working conditions and of the way BOOL_1 was developed, are crucial steps to understand how other processes of analysis would be built later. While the whole procedure is described in paragraph 6.3.1, in this section I describe the principal features that characterized this explorative analysis.

#### *5.5.1.1. A single behavioral pattern determined by three different profiles of activity*

The first run of the DCI algorithm revealed a result that has to be considered crucial for all the subsequent DCI procedures that have been carried out in the context of this work. A large group of agents turned out to be the best in terms of DCI performance. This first mask is made up of 55 agents and the other top 11 masks are very similar in terms of $t_{ci}$. The union of these 12 overlapping masks is determined by a set of 57 agents, with a minimum number of 53, a maximum number of 55, and an intersection of 45. Observing their profile of activity over time, it was immediately noticed that the actors involved had three different profiles of behaviors (see figure 5.9b):

- most of them participated only in the first part of the considered period;
- a part of them participated only in the last period;
- a small subset of them participated both at the beginning and at the end.

---

[81] A boolean variable takes value 0 if the agent is not active, and it takes value 1 if the agent is active. In this analysis 59 variables were elaborated, one for each identified instant in time. The activity of agents has been defined as the participation to at least one funded project.

[82] The software is loaded only on computers that are in the Department of Economics 'Marco Biagi' and at that time it was not possible to control them remotely. This caused the fact that every time a run finished, I had to re-launch it on the terminal.

It is interesting to notice that in this union set emerges a very particular pattern of activity: the coexistence of three different schemes of behavior. This result enhanced the interest in DCI analysis, but at the same time some problematic aspects were revealed, and they have to be investigated further to be able to manage the whole process better.

### *5.5.1.2. Similar masks at the top of the $t_{ci}$ ranking list*

The first problematic aspect that emerged regarded the detection of a non-arbitrary principle to select the resulting masks. In this first run, having 12 masks at the top of the ranking every one very similar to each other, created some difficulties in deciding how to manage these results: should only the best mask be considered? Or should the union set be considered? Or the intersection set? Should the group of agents that occurred most frequently in these masks be considered? As can be easily understood, every possible criteria of interpretation of results involves some arbitrary evaluation. Moreover, there were two other important aspects that have to be considered:

- can a single run of the algorithm be deemed to be sufficient to obtain reliable results?
- the problem of managing the top masks detected by the algorithm has been described, but what about all the other masks? How is it possible to get an overall interpretation of results?

Reflections about these aspects were determinant in the continuation of the analysis. Especially after the upgrade of the software that allowed DCI the algorithm to run automatic repetitions, it was possible to easily overcome the problem of a sufficient number of repetitions of the analysis.

*Figure 5.9a (on top). Best 12 subsets detected in the first round of BOOL_1 model. Every column represents an agent and every row represents a different subset. A black cell indicates that the corresponding agent belongs to the corresponding mask. Masks are in descending order of $t_{ci}$ (the best on top). Only agents involved in these masks were put in the figure. It is possible to observer that these masks are very similar, but they are not identical. Above the masks there is row in which are written the identification number that those agents had in the first round of the BOOL_1 model. Below the masks there is row in which are counted the number of masks in which every agent is present.*

*Figure 5.9b (under figure 5.9a). The profile of activity of the agents involved in the masks of the figure on top. Every column represents an agent and every row represents a different instant in time (in total they are 59 and they are ordered in chronological order from the top). A red cell indicates that the corresponding agent was active in the corresponding instant. It can be observed that the behavioral pattern that was detected is made up by three different kinds of profile of activity: (i) agents that participated in the first period (left); (ii) agents that participated in the second part of the policies (middle-right); (iii) agents that participated in both (right).*

On the other hand, the necessity that emerged to identify a reliable solution to group the masks, led to the introduction of a process of cluster analysis, as described in paragraph 5.4. With this solution both the problem of how to consider similarity between masks and the problem of having an overall interpretation of results, were solved.

### *5.5.1.3. Economic interpretation of the first identified group*

Methodological issues have already been explained, but still what can be considered the most important result from an economic perspective has to be illustrated. In fact, in addition to all previous considerations, it is fundamental to report the fact that the group of 57 agents that belong to the union set of the top 12 masks of the first run of BOOL_1, is a group of agents that have very little importance from an economic point of view. A large part of them (19 out of 57) is composed by local public institutions of small territories with a small number of participations, and the other big portion (22 out of 57) is made up of enterprises that participated in only two funded projects (the minimum imposed level of participation). So, it can be said that the first run of DCI brought to light actors that seem not to have had a key role in the context of considered policies. Looking also at the profiles of activity of these agents, the impression is that they were immediately collocated in the group with the highest $t_{ci}$ score because of the fact that the few participations they had, were extremely integrated.
Of course, it was not expected that a group would be detected in which three different schemes of activity coexist to determine a unique behavioral pattern[83]. Thus, it cannot be denied that a very interesting result was immediately encountered but at the same time the conviction was strong that the analysis had not yet reached the core. The intention was to make DCI able to return information concerning those agents that in the considered context have a stronger importance from the perspective of regional economy and innovation processes.

### *5.5.1.4. Skimming the initial set and iterating the analysis: rounds of DCI algorithm*

These considerations shed new light on the first result obtained with DCI analysis. In fact, even if DCI was not designed to suggest agents that should be excluded, it seemed evident that the 57 agents included in the 12 masks with the highest $t_{ci}$, should have been ruled out from the analysis before starting it. Because of their uninteresting role and because of their low level of activity, these agents can hardly represent the core of a community structure devoted to innovation. In other words, these agents should have been skimmed from the considered system, but there was no way to do it without arbitrary evaluations.
As explained above (see paragraph 5.3.3), the reduction of the number of agents involved in the analysis, led to the consideration of those 352 agents with at least two participations in funded projects. Starting from 1.121 agents in total, the exclusion was intense but it could go further: the decision that prevented an ulterior cut in the system that would definitively be considered for the analysis concerned the fact that it would have caused a too high a loss of enterprises. At the same time, it was not fair to preliminarily exclude agents on the basis of their institutional typologies and, even if small local government institutions do not generally

[83] See paragraph 5.5.1.1

represent a category of very interesting actors, there are some that maintain a crucial role. For these reasons, the option to repeat the DCI analysis over the initial system (352 agents) excluding the group of agents identified through the described 12 masks (57 agents) seemed to be meaningful and appropriate and so, after the first round[84], a new round of the DCI analysis was launched over a set of 295 agents.

#### *5.5.1.5. Main features of the process BOOL_1*

Without describing here all the rounds and all the results that characterize the process BOOL_1, it is necessary to briefly summarize the main features of this model of analysis:

- initial set made up of those 352 agents that participated in at least two funded projects;
- progressive skimming of the most relevant masks (in terms of $t_{ci}$) detected at the end of every round;
- variable criteria in selecting the most relevant groups of agents that emerged from every round;
- variable number of runs of the algorithm in every round;
- 25 generations produced by the Genetic Algorithm for the analysis of every considered dimension;
- variable maximum dimension of considered subsets[85], due to hardware computational limits;
- every group identified in every consecutive round was defined as a subsystem, a *modus operandi* which allowed me to focus attention progressively on those agents that showed most complex behavioral patterns but which prevented me from finding overlapping communities (except two particular cases[86]);
- after eight progressive and consecutive skimmings, in a system that had reached the dimension of 92, a subset of 73 agents was detected as the most relevant group in terms of $t_{ci}$. The decision was to stop the skimming procedure and the ninth round of the analysis continued on these 73 agents instead of restricting the analysis to their complementary subset;
- after two rounds (the ninth and the tenth) in which the DCI algorithm was launched on the group of agents that emerged from their previous round, in the last round of the analysis (the eleventh) was considered a system made up of (i) those agents that were not considered at the end of the eighth round (19 agents) plus (ii) those agents that were not considered at the end of the ninth round (18 agents) plus (iii) those agents that were not in the best group detected at the end of the tenth round (21 agents). This fact permitted the identification of a last subsystem (made up of 41 agents) that overlaps with some of the subsystems identified in previous rounds of the analysis.

---

[84] The first round was determined by a single run, since no other runs were made.

[85] To let the algorithm run more than one analysis simultaneously, it was necessary to reduce the maximum dimension of considered subsets.

[86] In round 6 and in round 7 the masks with the highest score in terms of $t_{ci}$ seemed to depict an overlapping structures made by different subsystems.

### *5.5.1.6. Final observations on process BOOL_1*

As can be observed, BOOL_1 was a process that represented a completely new exploration. When the analysis started, there was no idea about what the results would be, and maybe what is more interesting to remember, is that there was no idea of what the difficulties would be and what problems would have to be faced. All these elements provided challenges made up of new aspects that have to be debated and managed.

The results of BOOL_1 is discussed in more detail later. Here it is important to highlight some considerations:

- the problem of having an incredibly high level of similarity among the masks with the highest level of $t_{ci}$, was tackled with a meditated and arbitrary process of selection of some kind of set of them (union, intersection, the group of the most frequent agents). Of course, the necessity emerged to find a less arbitrary criterion, but at the same time the procedure was carried out with the greatest attention to the step-by-step observation of the emerging groups;
- the variability in the number of times that the algorithm was launched (number of runs) in every round of the analysis should have been fixed with more attention, but at that stage of the work computational limits constituted a serious obstacle that caused severe time dilations. After a fundamental upgrade of the software it was possible to sensibly reduce the time taken by a single run, and it was possible to set the automatic iteration of runs and the final overall ranking list of masks. This reduced the possibility that the algorithm, in particular because of the initial random sampling of subsets of minimum dimension, could embark on a wrong path;
- the discovery that groups (especially in first rounds) with the highest levels of $t_{ci}$ were groups of uninteresting actors (regarding a context of analysis of regional economic processes of innovation), led to the necessity to guide the analysis in the direction of those agents that for many reasons were considered as holders of key roles. The solution that seemed to be most appropriate, was an iteration of the analysis over subsets determined by a progressive skimming of the detected groups. In this context arose the idea of doing progressive rounds of the analysis;
- even if during the whole process maximal attention was paid to taking account of groups' composition, the necessity to establish a more solid and non-arbitrary criterion in the managing of the rounds emerged. The most important theoretical and methodological problem is that every time a group is skimmed from the whole, the system under analysis changes. From a theoretical perspective, it should be noted that a community is defined not only for its internal dynamics but also for the relation that it has with the rest of the environment. Here the environment is composed of the entire set of agents that participated at least twice in regional public policies and when a part of it is excluded before the beginning of a round, a completely new and diverse system is produced. As long as we accept that the definition of a community necessarily involves the comprehension of its relation with the rest of the world that it belongs to, it is not appropriate to reconstruct the presence of communities through a progressive skimming of the system under observation. In fact, de-stratified subsets from the entire initial set have to be considered as diverse systems with their own dynamics. Moreover, from a

methodological point of view, it must be remembered that in no circumstances can the $t_{ci}$ of subsystems identified over different systems be compared. This fact led to the conviction that the process of skimming should be used to allow the DCI analysis to be launched over the most interesting group of agents. However, after a point that the researcher has to identify, the de-stratification must be stopped and the detection of different overlapping communities must be done over a unique system.

### 5.5.2. Process of analysis 'BOOL_2'

The second process of analysis with DCI methodology, was done using the same informational basis as BOOL_1: 352 agents and 59 boolean variables that describe the activity status (in a binary form) of agents over time, with respect to their participations in regional public policies. All considerations and reflections that emerged from the first process of analysis BOOL_1, led to the desire to elaborate another model that would be less arbitrary. The choice was to conduct a new analysis that, starting from the same initial basis as the previous one, could advance trying to overcome the perplexing aspects that emerged during the first process. Because of that, this new process is characterized by a more rigid structure that tries to furnish a more reliable support for the revealed difficulties. This new process of analysis identifies a completely new model, and it will be referred to it as 'BOOL_2'. Here the main features that characterize the model are summarized:

- initial system determined by those 352 agents that participated in at least two funded projects in the context of considered regional public policies;
- 59 boolean variables describing in a binary way the status of activity of each agent over the 59 detected instants in time;
- minimum dimension of evaluated subsets equal to 2;
- maximum dimension of evaluated subsets equal to 90;

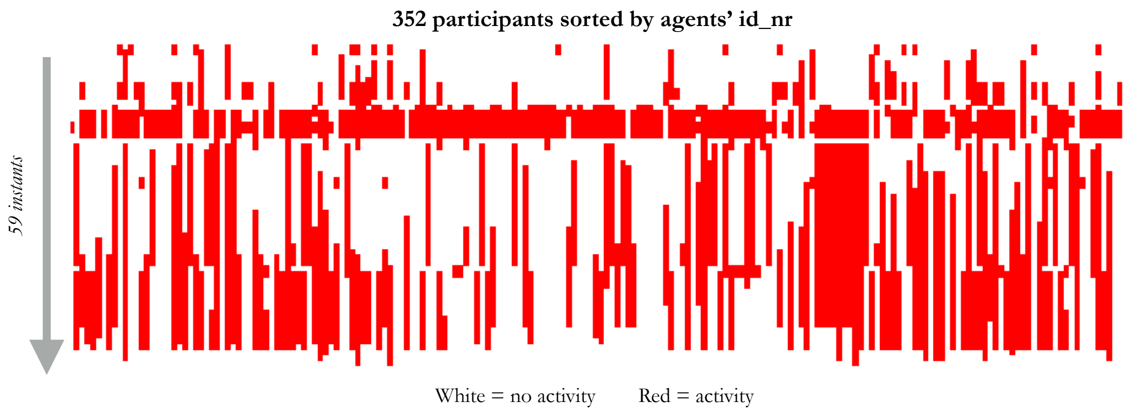



*Figure 5.10. The boolean profile of activity of the 352 agents considered in the analysis. Every column represents an agent (sorted by identification number) and every row represents a different instant in time (in total they are 59 and they are ordered in chronological order from the top). A red cell indicates that the corresponding agent was active in the corresponding instant. If the cell is white, the agent was not active in that particular instant.*

- 25 generations produced by the Genetic Algorithm for the analysis of every considered dimension;
- selection of the mask with the highest $t_{ci}$ value at the end of every round;
- iteration of the analysis over several rounds;
- progressive skimming, round after round, of the mask selected at the end of the previous round;
- 30 runs of the algorithm in every round and single final ranking of all analyzed masks;
- stopping of the skimming procedure after the substantial loss of agents that the researcher retains must not be excluded;
- hierarchical agglomerative cluster analysis of 12.500 masks with the highest $t_{ci}$ value: simple matching as similarity criterion and complete linkage as agglomerative method.
- dendrogram cut with respect to the length of branches, to the identification of clusters with a good degree of diversity among them, and to the involvement of a good number of agents;
- in every cluster, selection of the mask with the most significant score of DCI, as the representative mask of that cluster.

### 5.5.3. Process of analysis '∂LoA'

After exploring the process of analysis BOOL_1 the decision was taken to experiment other models with an ameliorated structure in terms of methodology. This is what has been done with BOOL_2 and with another process of analysis which will be called '∂LoA'. The acronym can be read in this way: '∂' stands for 'variations' and 'LoA' stands for 'Levels of Activity'. The procedure is exactly the same as that of BOOL_1, but there is an important change in the considered informational basis. Instead of considering boolean variables, here variables were used that express the relation between the number of projects that were active in a specific instant, and number of projects that were active in the previous instant. In other words, the number of funded projects that every agent was participating in every one of the 59 instants in time was observed and, after having maintained these instants in a chronological order, 58 variables were elaborated in this way:

- if the agent has a number of active projects higher than the number of projects the agent was participating in, in the previous instant, the corresponding variable assumes value 3;
- if the agent has a number of active projects equal to the number of projects the agent was participating in, in the previous instant, and the number is greater than zero, the corresponding variable assumes value 2;
- if the agent has a number of active projects lower than the number of projects the agent was participating in, in the previous instant, and the number is greater than zero the corresponding variable assumes value 1;
- if the agent is participating in no funded projects, it has to be considered inactive and so the corresponding variable assumes value 0.

Obviously, considering variation in time implies the reduction of the number of variables that in fact passes from 59 to 58. The information about the activity/inactivity of agents does not change, since if an agent is participating in no funded projects the variable continues to assume value zero, as in boolean models. What was done was a further step in the characterization of the activity status. In particular, it seemed to be very interesting to investigate the integration among agents with respect to the increase (or decrease) of the number of their participations. This is a very interesting perspective for two reasons:

- what is observed is not only how groups of agent show behavioral patterns in terms of presence of activity (or not), but also how the number of funded projects of every agent increases or decreases in relation to what other agents do. In this way what is investigated is a level of integration of behaviors that is finer than boolean variables allow, since it can be observed if the activity is increasing, is remaining equal, is decreasing or is null;
- the elaboration of this new series of variables allows the real introduction of the time dimension. While with boolean variables the time was considered only through the necessary identification of instants in which to observe agents, in this new model the activity of agents is described in terms of evolution. Of course, the level of description refers to only one time lag, but this new perspective is fundamental to include in the analysis the time dimension as a crucial element that characterizes the agents' dynamics and, therefore, the process of evolution of the whole system and of its communities.

All features of model ∂LoA except the nature of variables (variation of levels of activity instead of boolean status), are exactly the same as BOOL_2, as they are described in the previous section.

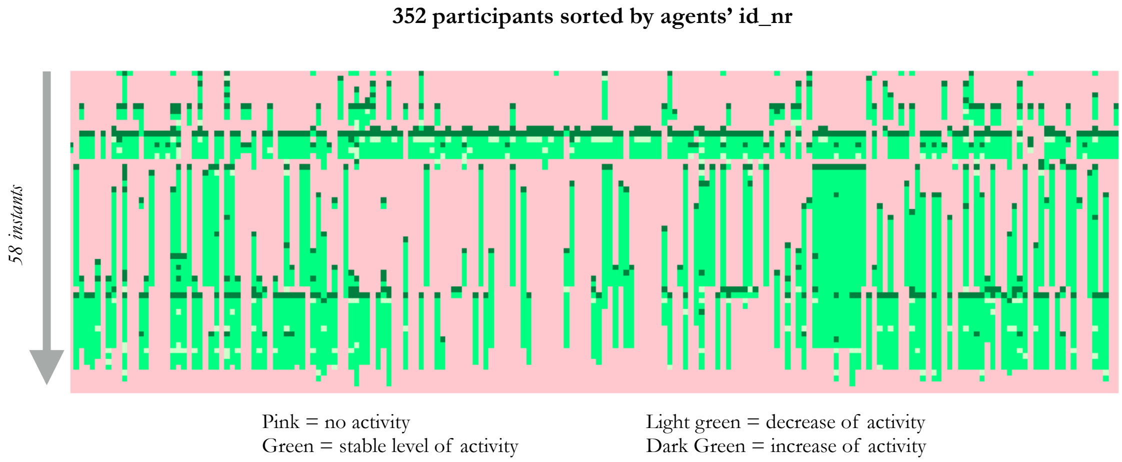



*Figure 5.11. The profile of activity (variation of levels of activity) of the 352 agents considered in the analysis. Every column represents an agent (sorted by identification number) and rows represent instants in time ordered in chronological order from the top (here they are 58 because the status of activity is calculated as a comparison between the number of projects in which the agent was participating in a specific instant and the number of projects in which the agent was participating in the previous instant, so the first instant is not present). A pink cell indicates that the agent was not participating in any funded project. A light green cell indicates that the agent (in the corresponding instant) was participating in a number of projects lower than the number of projects in which it was participating in the previous instant. A green cell indicates that the agent (in the corresponding instant) was participating in a number of projects equal to the number of projects in which it was participating in the previous instant. A dark green cell indicates that the agent (in the corresponding instant) was participating in a number of projects higher than the number of projects in which it was participating in the previous instant.*

# 6. Results

The three methodologies that were taken into account were set according to the different considerations explained in chapters 3, 4 and 5. For each methodology, three models were developed with regard to (i) the theoretical accordance between the available applicative options and their theoretical implications, (ii) the theoretical interest in a more specific tailoring of the methodology, and (iii) the manageability of each methodology with respect to the possibility of setting it computationally. As was described in the dedicated chapters, each model proposed has a particular meaning with respect to the case study. The conceptualization and the implementation of these aspects should be considered the core of this work.

After the conceptualization of the different models, their applications produced nine different partitions of the considered set of agents (352 agents that participated in at least two funded projects). In this chapter, a description of these partitions, and of the characteristics of the communities that belong to each one of them, are discussed. Considering one by one the three methodologies, every partition produced will be described in its essential and most interesting features. The descriptions of the detected partitions are developed on two main levels. In the first, more focused on the shapes of the partitions, the structure of the communities and the connections that occur among them are analyzed. In the second, more centered on the specific features of the identified communities, attention is given to contingency tables regarding some features (in particular, the participations in different waves and the participations in projects with specific technological domains) of the agents involved. In order to highlight specific characterizations of these communities, the adjusted residuals[87] of the corresponding contingency tables are considered.

Finally, in order to understand if there are groups of agents that simultaneously respond to the aspects investigated with the different methodologies, I evaluate the Jaccard similarity coefficient[88] of all possible combinations made up of three communities as follows:

- the first community is one of the communities detected with CPM;

[87] Adjusted residuals, defined by Haberman (1973), are defined as the raw residuals (or the difference between the observed counts and expected counts) divided by the estimated standard deviation of the observed count. Formally:

$$d_{i,j} = e_{i,j} / \sqrt{[(1 - n_{i,+}/N)(1 - n_{+,j}/N)}$$

where ei,j is the standard error (difference between the observed frequency and the expected frequency, divided by the root of the expected frequency) and (ni,+/N) and (n+,j /N) are respectively the row proportion and the column proportion.

[88] The Jaccard similarity coefficient is a coefficient of similarity for binary variables. Indicating with letters *a, b, c, d* the four possible combinations of two binary variable, and indicating with capital letter the frequency that the corresponding situation occurs between variables of *i-th* observation and variables of *j-th* observation, it is possible to define the Jaccard similarity coefficient as

Jaccard similarity coefficient = A / (A + B + C) , where:

a=(1,1) b=(1,0) c=(0,1) d=(0,0)

Considering every community as a binary variable that assumes value 1 if the corresponding node belongs to the community and 0 if it does not, the Jaccard coefficient represent the ration between the dimension of the intersection and the dimension of the union set of the couple of communities considered.

- the second community is one of the communities detected with Infomap;
- the third community is one of the communities detected with DCI.

In doing this, I will make further considerations about the overlaps among communities that were detected with methodologies which, moving from different perspectives, should put in evidence specific topics of investigation.

## 6.1. CPM communities

The three models developed in the context of CPM differ from each other for the value of *k*. As described in chapter 3, the selection of a specific value could not find any support from the innovation theory and, for this reason, when I joined the working group, the development of CPM analysis was in a state of stasis. The decision regarding the selection of specific values of *k* was taken thanks to a shift in the perspective adopted. Typically, the value of *k* is defined on the basis of considerations regarding the core network structure that has to be investigated: the clique. Unfortunately, in this specific case no particular suggestions were present. In such a context, some attempts were made to select the value of *k* on the basis of some information about communities' dimensions, but the situation was still confused since the only observations that were made regarded:

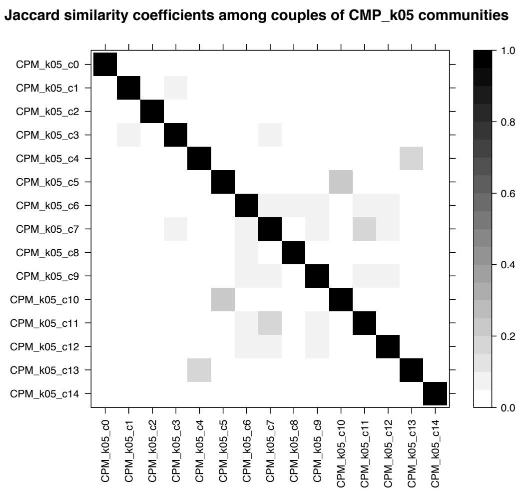
- the decrease in the number of agents that belong to at least one community;
- the decrease in the standard deviation of the dimension of the communities detected for every value of *k*;
- the increase and the subsequent decrease in the number of communities detected.

As shown in the third chapter, considerations regarding these aspects did not provide sufficient elements. Then, to overcome this difficulties, I proposed to focus the attention on other kinds of considerations and what seemed to me determinant was to observe the structure that every partition produced. In particular, the presence of agents that belong to more than one community revealed different degrees of overlaps in different partitions. The connections that were found among communities - two communities are connected if they have at least one node in common, i.e. if they overlap - made it possible to understand how detected partitions are different. The observation of how communities overlap was determinant to separate the possible values of *k* in three ranges. The first range ($3 \leq k \leq 8$) is characterized by partitions with an increasing number of communities (as *k* increases), a high standard deviation of the dimensions of the communities, and a not too large number of connections among communities. The second range ($9 \leq k \leq 16$) identifies partitions with a high number of communities, a large number of connections and a high homogeneity in terms of communities' dimensions. The last range ($17 \leq k \leq 22$) is characterized by partitions with a small number of communities, the same degree of homogeneity that was observed in the second range, and few connections among communities. At the end, in every one of these three ranges the value of *k* that seemed to me the most interesting and representative

was selected. In this way, three models were developed using respectively *k* equal to 3 (model CPM_k05), *k* equal to 12 (model CPM_k12) and *k* equal to 18 (CPM_k18).
The three partitions obtained are described one at the time. It is important to underline that, while with other methodologies the partitions detected have to be investigated from the perspective of their structure, the CPM models were expressly developed thanks to the observation of the structures themselves. Thus, here only some elements regarding how the communities are connected among themselves are taken into account and most attention is directed to the observation of other aspects that can allow to highlight the most important features of the three considered models.

### 6.1.1. CPM_k05

The model CPM_k05 depicts a partition completely dominated by the presence of a huge community (C2) to which belong the 89% of the 334 nodes that are included in at least one community, while other communities have an average dimension of 7,57 nodes (with a standard deviation equal to 3,72). After the identification of the intersections among all possible couples of communities, I calculated the Jaccard coefficients of similarity between all pairs of communities (figure 6.1) but this was not sufficient to understand what the pattern of the overlaps is like.



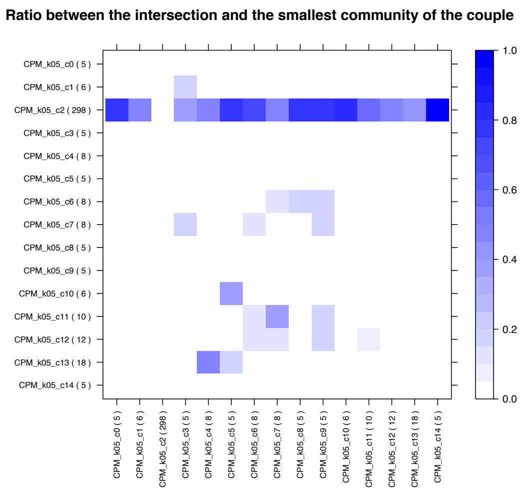




*Figure 6.1 (left). Matrix of Jaccard similarity coefficients calculated for all pairs of model CPM_k05 communities. The darker is the color, the higher is the ratio. Rows and columns represent (ordered from the top to the bottom and from left to right) CPM_k05 communities.*

*Figure 6.2 (right). Matrix of ratios between the dimension of the intersection and the dimension of the smallest community (in the considered couple) of all pairs of communities detected in model CPM_k05. A ratio equal to 1 (dark blue) identifies combinations in which the smallest community is completely overlapped to the other community. If the ratio is equal to 0, no overlaps are present. The main diagonal has been intentionally set equal to 0 to allow a clearer vision of the figure. Rows and columns represent (ordered from the top to the bottom and from left to right) CPM_k05 communities.*

To have a better understanding of it, I computed the ratios between the dimension of the intersection and the dimension of the smallest community of all the possible couples[89]. It emerged clearly that all communities are intensively connected with the biggest one, as it is possible to observe in figure 6.2. Observing the combinations with the most intense color (that correspond to ratios close to 1), it can be immediately understood that all the communities share a big part of their nodes with the biggest community. Since this partition depicts a situation where a huge community has intense overlaps with all the other communities, it seemed to me that it can with difficulty suggest interesting elements of what I was looking for: a relational structure through which agents interacted over the whole cycle of the considered public policies. The presence of this overwhelming group of agents necessarily could not allow those more detailed parts of the framework of connections that I was investigating to emerge.

### 6.1.2. CPM_k12

As said above, the partition generated by model CPM_k12 is characterized by the presence of a large number of communities (27) with a high degree of overlapping (average number of connections per community equal to 7,40) and a homogeneous distribution of the number of nodes that belong to every community (average equal to 17,62 and standard deviation equal to 6,55). While the preceding partition (CPM_k05) cannot suggest important elements because of the presence of a single community that overshadows the entire network, here the difficulty to provide a meaningful interpretation is due to the presence of a such large number of connections that inevitably confuses the interpretation of the results. The presence of 200 overlaps, which on average include 2,29 agents, clearly depicts a partition in which all the communities have a large number of connections among each other, and in which these connections are weak. In this context, the investigation of the structure is difficult since the homogeneity of the dimension of the communities, along with a dissemination of the overlaps, gives rise to a fuzzy partition in which it is difficult to glimpse relevant elements. However, some considerations can be made thanks to the observation of the contingency table describing how agents involved participated in the different waves.

Rather than take into account the observed frequencies of agents' participations in the different waves, what it is important to do is to focus the attention on the corresponding adjusted residuals. In table 6.5, the cells that have an adjusted residual higher than or equal to 5 have been colored in green and immediately the presence of several cells with such high values can be

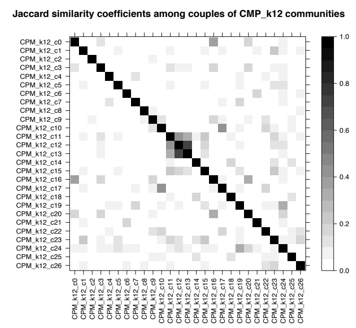



*Figure 6.3. Matrix of Jaccard similarity coefficients calculated for all pairs of model CPM_k12 communities. The darker is the color, the higher is the ratio. Rows and columns represent (ordered from the top to the bottom and from left to right) CPM_k12 communities.*

[89] The ratio between the dimension of the intersection and the dimension of the smallest community in the considered couple, obviously is included between 0 and 1. If the ratio is 0, it means that the intersection is null, while a ratio equal to 1 reveal a situation in which the smallest community is completely included in the biggest one.

| Obs. Freq. & Adj. Res. | Waves | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| CPM_k12 Comm. | 2002_171 | 2002_172 | 2002_IT T | 2004_171 | 2004_171 A | 2005_171 | 2006_VI N | 2007_171 |
| C0 | | | 71 | | | 3 | 4 | |
| C1 | 11 | 10 | 45 | | 14 | 24 | | |
| C2 | 3 | 8 | | | | 37 | | |
| C3 | | | 77 | | | | | |
| C4 | 8 | | 24 | | 8 | 37 | | |
| C5 | | 5 | 35 | | 9 | 28 | | |
| C6 | 8 | 8 | | | | 38 | | |
| C7 | | 5 | 34 | | | 46 | | |
| C8 | 3 | 6 | 6 | 4 | | 38 | | |
| C9 | 22 | 10 | 14 | 2 | 17 | 28 | | |
| C10 | 30 | | 13 | | | 62 | | |
| C11 | 12 | 5 | 32 | | 15 | 72 | | |
| C12 | | | 12 | | 6 | 49 | | |
| C13 | | | 15 | | 6 | 22 | | 2 |
| C14 | 39 | 12 | 42 | 27 | | 33 | | 8 |
| C15 | | 7 | 8 | 8 | 15 | 64 | | |
| C16 | | | 79 | | | 16 | | |
| C17 | 19 | 4 | 30 | | | 79 | | |
| C18 | 31 | | 2 | 33 | | 19 | | |
| C19 | | 9 | 30 | 5 | | 67 | | |
| C20 | | | 87 | | | | | |
| C21 | 5 | | 25 | | 4 | 31 | | |
| C22 | 38 | 8 | 26 | | 7 | 109 | | |
| C23 | 18 | 10 | 50 | | 15 | 56 | | |
| C24 | | 13 | 65 | 5 | | 81 | | |
| C25 | 117 | | | | | | | |
| C26 | 5 | | 13 | 4 | 9 | 43 | | |

| Waves | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 2002_171 | 2002_172 | 2002_IT T | 2004_171 | 2004_171 A | 2005_171 | 2006_VI N | 2007_171 | Number of cells with Adj. Res. ≥ 5 by row |
| -3,62 | -1,96 | 11,43 | -1,67 | -2 | -6,79 | 11,46 | -0,55 | 2 |
| -1,03 | 2,52 | 2,58 | -1,94 | 4,26 | -3,81 | -0,41 | -0,64 | |
| -1,56 | 4,06 | -4,77 | -1,3 | -1,56 | 5,12 | -0,27 | -0,43 | 1 |
| -3,6 | -1,95 | 13,07 | -1,66 | -1,99 | -7,44 | -0,35 | -0,55 | 1 |
| -0,93 | -1,95 | -0,1 | -1,66 | 2,36 | 1,26 | -0,35 | -0,55 | |
| -3,6 | 0,83 | 2,63 | -1,66 | 2,91 | -0,86 | -0,35 | -0,55 | |
| 0,17 | 3,65 | -5,06 | -1,38 | -1,66 | 4,42 | -0,29 | -0,46 | |
| -3,78 | 0,6 | 1,67 | -1,74 | -2,09 | 2,48 | -0,37 | -0,58 | |
| -1,92 | 2,18 | -3,48 | 1,56 | -1,7 | 3,97 | -0,3 | -0,47 | |
| 2,73 | 2,92 | -3,52 | -0,65 | 6,25 | -2,19 | -0,38 | -0,61 | 1 |
| 4,39 | -2,29 | -4,34 | -1,95 | -2,34 | 3,82 | -0,41 | -0,65 | |
| -1,79 | -0,51 | -2,11 | -2,23 | 3,54 | 2,88 | -0,47 | -0,74 | |
| -3,35 | -1,81 | -2,46 | -1,54 | 1,64 | 5,4 | -0,32 | -0,51 | 1 |
| -2,73 | -1,48 | 0,24 | -1,26 | 2,73 | 1,07 | -0,26 | 4,47 | |
| 3,85 | 1,82 | -1,58 | 9,78 | -2,92 | -5,48 | -0,51 | 9,77 | 2 |
| -4,16 | 1,14 | -5,28 | 2,58 | 4,82 | 4,53 | -0,4 | -0,64 | |
| -4,01 | -2,17 | 10,98 | -1,85 | -2,22 | -4,89 | -0,39 | -0,61 | 1 |
| 0,13 | -0,86 | -2,28 | -2,19 | -2,63 | 4,49 | -0,46 | -0,73 | |
| 6,06 | -2,05 | -5,91 | 18,5 | -2,09 | -3,57 | -0,37 | -0,58 | 2 |
| -4,35 | 1,83 | -1,08 | 0,7 | -2,4 | 4,22 | -0,42 | -0,67 | |
| -3,83 | -2,07 | 13,92 | -1,76 | -2,12 | -7,92 | -0,37 | -0,59 | 1 |
| -1,49 | -1,78 | 1,18 | -1,52 | 0,54 | 1,1 | -0,32 | -0,5 | |
| 2,54 | -0,21 | -5,47 | -2,65 | -0,69 | 4,88 | -0,56 | -0,88 | |
| -0,7 | 1,3 | 0,5 | -2,34 | 3,14 | -0,9 | -0,49 | -0,78 | |
| -5,34 | 2,14 | 2,25 | -0,22 | -2,95 | 2,23 | -0,52 | -0,82 | |
| 27,41 | -2,42 | -7,54 | -2,06 | -2,47 | -9,24 | -0,43 | -0,68 | 1 |
| -1,82 | -1,91 | -2,65 | 1 | 3,04 | 3,02 | -0,34 | -0,54 | |

*Table 6.4 (left). Contingency table regarding communities and policies waves. The cells contain the observed frequencies of the participations which those agents that belong to the corresponding CPM_k12 community (rows) had in the corresponding policy wave (columns). Since communities can overlap, a same agent can belong to more than one community. Thus, in every community to which an agent belongs, all its participations are considered.*

*Table 6.5 (right). Adjusted residuals of the Table 6.4. In green, adjusted residuals higher than or equal to 5. On the right, counting of the cells (by row) that have an adjusted residual higher than or equal to 5.*

noticed. Moreover, their distribution over the different communities is interesting. In fact, out of a total of 13 cells with an adjusted residual higher than or equal to 5 (which is a very high threshold), there are no communities with more than two cells with values of this magnitude. The characterization in terms of participations in the waves is fairly uniformly distributed and does not present communities with particular peaks in terms of number of significant combinations. This result, which may for the moment be seen with a limited interest, will have to be remembered later, after the reading of the results of model CPM_k18.

### 6.1.3. CPM_k18 and comparison of the three CPM models

The last CPM model is the one with the value of *k* equal to 18. The increase in the value of *k* implies that the dimension of the fundamental clique that is 'rolled' over the network has reached an order of magnitude that is close to its maximum (22). The capacity of the clique to move over the network is affected by its big size and the number of agents that belong to

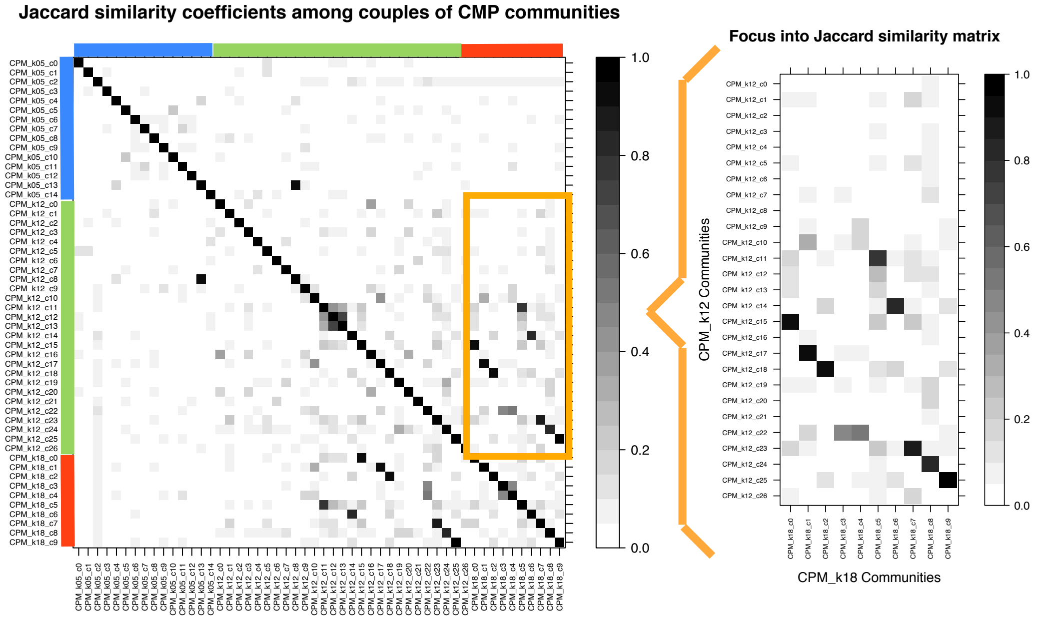



*Figure 6.6a (left). Matrix of Jaccard similarity coefficients calculated for all pairs of communities detected in all three CPM models (CMP_k05 highlighted in blue, CPM_k12 in green and CPM_k18 in red). The darker is the color, the higher is the ratio. Rows and columns represent (ordered from the top to the bottom and from left to right) CPM_k05, CPM_k12 and CPM_k18 communities.*

*Figure 6.6b (right). Particular of Figure 6.6a. Jaccard similarity coefficients between CMP_k12 communities (rows) and CPM_k18 communities (columns). The darker is the color, the higher is the ratio between the intersection and the union set of the pair of communities considered.*

at least one community confirms this limited capacity. While with CPM_k05 there were 334 agents involved[90] and while with CPM_k12 this number was 250, with CPM_k18 there are 142 nodes that are present in the community structure detected by the partition. Thus, we are considering a number of agents that is smaller than 50% of the initial set (that is made up of 352 agents) but, instead of interpreting this situation as a problematic aspect, it can be regarded as an element of interest. In fact, since the number of involved agents is small because of the difficulty of the clique to move over the network, it is to be expected that the fuzzy situation that was observed in model CPM_k12 should crystallized.

This is what happened. In fact, observing the Jaccard similarity matrix computed with all possible combinations of communities belonging to all three CPM models (figure 6.6a), it is possible to notice a peculiarity in a specific portion of it. If the attention is focused on the combinations between CPM_k12 communities and CPM_k18 communities, as represented in figure 6.6b, it appears very clearly how every column has at least one cell in which the considered similarity coefficient is close to 1 (the cells with a ratio close to 1 are colored in dark grey and they are black if the ratio is equal to 1). This is interesting because it depicts a situation in which every CPM_k18 community has an intense intersection (from the point of view of the ratio between intersection and union set) with one of the CPM_k12 communities. Moving into a value of *k* equal to 18 has produced the loss of some of the groups detected with *k* equal to 12. That is the reason why I spoke about a crystallization,

[90] Involved agents are those that belong to at least to one community. It has to be reminded that the application of the considered methodologies does not imply that all the considered agents (352) have to belong to the system depicted by the corresponding partition. Agents can stay out of every detected community.

| Obs. Freq. | Waves | | | | | |
|---|---|---|---|---|---|---|
| **CPM_k18 Comm.** | **2002 _171** | **2002 _172** | **2002 _IT T** | **2004 _171** | **2004 _171 A** | **2005 _171** |
| C0 | | 7 | 8 | 8 | 15 | 46 |
| C1 | | 4 | 30 | | | 74 |
| C2 | 24 | | 2 | 29 | | 19 |
| C3 | | 8 | 7 | | 7 | 64 |
| C4 | 23 | 8 | 19 | | 7 | 49 |
| C5 | 12 | 5 | 30 | | 15 | 60 |
| C6 | 28 | 12 | 42 | 27 | | 31 |
| C7 | 11 | 10 | 32 | | 15 | 56 |
| C8 | | 8 | 65 | | | 74 |
| C9 | 117 | | | | | |

| Adj. Res. | Waves | | | | | | |
|---|---|---|---|---|---|---|---|
| **CPM_k18 Comm.** | **2002 _171** | **2002 _172** | **2002 _IT T** | **2004 _171** | **2004 _171 A** | **2005 _171** | Number of cells with Adj. Res. ≥ 5 by row |
| C0 | -4,68 | 1,14 | -2,73 | 1,53 | 5,32 | 2,33 | 1 |
| C1 | -5,37 | -0,90 | 1,76 | -2,71 | -2,59 | 5,71 | 1 |
| C2 | 2,93 | -2,17 | -4,03 | 12,75 | -2,11 | -3,06 | 1 |
| C3 | -4,74 | 1,56 | -3,09 | -2,39 | 1,21 | 6,19 | 1 |
| C4 | 0,63 | 0,92 | -0,87 | -2,68 | 0,62 | 0,77 | |
| C5 | -2,83 | -0,76 | 0,97 | -2,90 | 3,64 | 1,54 | |
| C6 | 0,19 | 1,64 | 2,72 | 7,33 | -3,00 | -5,26 | 1 |
| C7 | -3,15 | 1,27 | 1,33 | -2,93 | 3,56 | 0,59 | |
| C8 | -6,39 | -0,09 | 7,33 | -3,22 | -3,09 | 2,01 | 1 |
| C9 | 23,31 | -2,79 | -5,93 | -2,83 | -2,71 | -9,87 | 1 |

*Table 6.7a (left). Contingency table regarding communities and policies waves. The cells contain the observed frequencies of the participations which those agents that belong to the corresponding CPM_k18 community (rows) had in the corresponding policy wave (columns). Since communities can overlap, a same agent can belong to more than one community. Thus, in every community to which an agent belongs, all its participations are considered.*

*Table 6.7b (right). Adjusted residuals of the Table X. In green, adjusted residuals higher than or equal to 5. On the right, counting of the cells (by row) with adjusted residual higher than or equal to 5.*

and a confirmation of this comes from the consideration of the structure of the partition. The communities detected in CPM_k18 model show a lower degree of overlapping but with a peculiarity. In fact, while the average number of connections (overlaps) per community decreases to a value of 2,60 (in the CPM_k12 partition it was 7,40), here the average number of agents in an overlap increases to 3,03 (in the CPM_k12 partition it was 2,29). Even if this variation is not huge, it is very interesting because it depicts a situation in which communities are less connected but the connections are more dense. The impression is that the fuzzy situation observed in the CPM_k12 partition has rarefied, making it possible now to glimpse those parts of the structure that seem to be characterized by bolder boundaries. This is exactly what I was looking for.
However, the most interesting results come from the observation of the adjusted residuals of the contingency table regarding the agents' participations in the different policy waves (table 6.7b). Even if in CPM_k12 an interesting distribution regarding high values of adjusted residuals was observed, in CPM_k18 this is clearer. In fact, it has to be noticed that, observing how adjusted residuals higher than or equal to 5 (colored in green in the table) are distributed over the communities, it can be noticed that there are 7 communities (out of the 10 detected) with a specific characterization regarding one specific wave of the considered policies. This is very interesting for two reasons. First of all, the characterization that was observed in CPM_k12 partition has acquired a clearer shape. This was not guaranteed, even if it is perfectly coherent with the observation of the intense intersections that every CPM_k18 community has with one of the CPM_k12 communities. Secondly, the observation of the specific characterization that every CPM_k18 community has with regard to the agents' participations in the different waves, can be regarded with interest because of the meaning that I assigned to the application of CPM to the specific case study.

| Number of communities with at least one adjusted residuals higher than or equal to 5 in the context of the corresponding contingency table | CPM_k05 | CPM_k12 | CPM_k18 |
|---|---|---|---|
| Typologies (agents) | 2 | 0 | 0 |
| Ateco 2002 Sottosezioni (agents) | 1 | 4 | 0 |
| Dimensional Classes (agents) | 0 | 0 | 0 |
| Waves (participations) | 12 | 10 | 7 |
| Technological domains (participations) | 10 | 18 | 8 |
| Number of communities in each model | 15 | 27 | 10 |

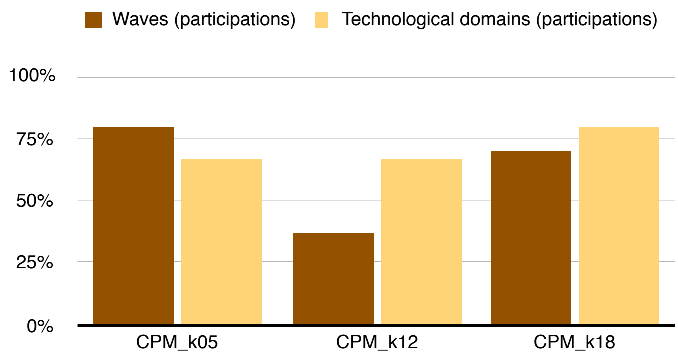



*Table 6.8 (above the chart). Counting of communities, by the model in which they were detected (column) with at least one combinations (in the contingency table with the corresponding category indicated in the rows) characterized by an adjusted residual higher than or equal to 5. In the first column, in bracket, is indicated if the category taken into account refers to attributes of the agents or of their participations. In the last row, the total number of communities detected in the corresponding model.*

*Chart 6.9 (above). Histogram of the percentages of the number of communities (over the total of communities detected in the corresponding model) that have at least one adjusted residual higher than or equal to 5 in the two different contingency tables that regard the participations in waves (brown) and the participations in projects with specific technological domains (light orange).*

Since I needed to investigate agents that could be grouped together because of the presence of a strong relational structure among them, CPM seemed to answered this necessity. In fact, it is interesting to observe that in the partition that highlights the essence of the structure (CPM_k18), the characterization of the communities concerns the agents' participations in the waves. This result seems to suggest that the structure of the network was determined first of all by the waves, and this is a very coherent hypothesis. If, on one hand, projects certainly determined intense relationships, on the other hand, waves determined the composition of the whole structure of edges. The possibility to participate more than once in the same wave (when allowed), along with the presence of some constraints, guided agents in the process of determination of which would become their adjacent nodes. This is the interpretation that I give to these results, and it is necessary to stress that a characterization geared towards agents' participations in waves was not obvious. There were no guarantees that the methodology (and the models) could permit the achievement of results that seem to be aligned with the hypotheses assumed. For these reasons, the last in particular, the application of CPM and the identification of a continuity among the three models has to be considered as an interesting development of the analysis. The first model, CPM_k05 identified a big community that, moving into other models, has rarified, allowing first of all the observation of fragmented partition affected by a large number of overlaps (in

CPM_k12), and then the identification of fewer but more distinct nuclei of agents (in CPM_k18).

To conclude the description of CPM models, a final consideration has to be made about another kind of contingency table, which will assume greater interest after viewing the results of Infomap models. What deserves attention is the presence of a high number of communities that in every CPM model has large adjusted residuals in the contingency table where the participations of agents are tabulated with regard of the technological domains of the corresponding projects. In this paragraph, particular attention was devoted to contingency tables regarding participations of involved agents in the different waves. Nevertheless, a very high number of communities with large adjusted residuals are observed also in the contingency tables regarding the technological domains of the projects in which agents participated (table 6.8 and chart 6.9). These results seem to show a degree of characterization higher than those regarding participations in waves. To have a better comprehension of this, I calculated the Jaccard similarity coefficient between every community detected in all CPM models and all the 168 partnerships of the 168 observed projects. These ratios were ordered in descending order and then plotted. In figure 6.10, it can be observed how different CPM models, in particular CPM_k12 (green circles) have communities which perfectly trace out the boundaries of some specific projects (if the community is equal to the partnership the ratio is equal to 1). Even if this seems to be a confirmation of the fact that CPM communities are characterized in terms of agents' participations in projects with specific technological domains, Infomap communities will show more interesting results. For this reason, the impression is that in the context of this analysis, CPM found its major significance under the profile of characterization concerning the agents' participations in different waves.

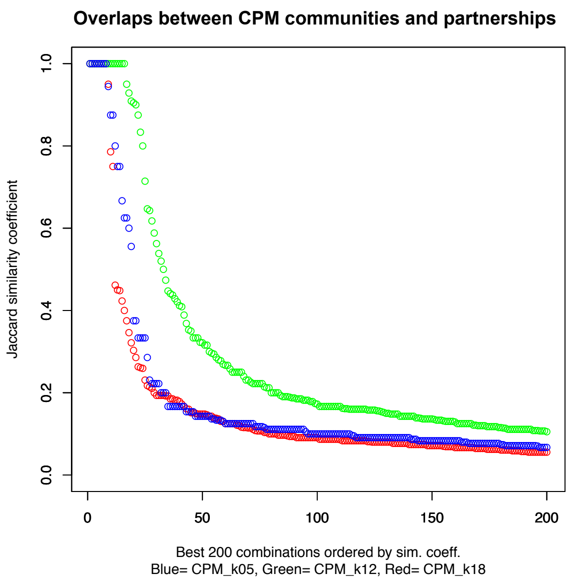



*Figure 6.10 (above). Scatterplot of the Jaccard similarity coefficients calculated between couple made up by (i) one of the communities detected in one of the CPM models (distinguished by different colors), and the partnerships configurations of the 168 projects that were realized in the context of the public policies. A coefficient equal to 1 indicates the perfect correspondence between the agents that belong to one of the communities detected and the agents involved in one of the observed projects. The calculated ratios were sorted in decreasing order and the best 200 coefficients were plotted. Every series has one circle for every position of the ranking (that is the X-axis). Thus, even if because of graphical computational limits does not appear so, the green circles with ratio equal to 1 do not start when the blue ones finish but they start in correspondence of value 1 of the X-axis.*

## 6.2. Infomap communities

The three models of Infomap methodology were developed according to different settings regarding the flow circulation over the network. The first model (MKV1) concerned the simulation of a memoryless flow that spread over the network without any kind of limit regarding the temporal sequence in which projects occurred. Since Infomap software includes also the possibility to apply second order Markov conditions, two other models were developed thanks to the reconstruction of a fake flow that took into account how projects came one after the other. Thinking of the simulated flow as a representation of the circulation of information, two different conditions made it possible to set two different models. In model MKV2_AS the possibility that the flow involved agents participating in simultaneous projects was allowed[91], while in model MKV2_DS the circulation of information was allowed only from terminated projects to projects that occurred later than the first ones[92]. From this perspective it can be said that, as happened in CPM models, a progressively more intense stress of the key parameter was imposed. While in CPM models this parameter regards the dimension of the clique, in Infomap models it concerns the application of more restrictive limits to the directions that the propagation of the flow can take. Considerations about the three models developed are presented hereunder and final observations are provided at the end of the paragraph.

### 6.2.1. MKV1

Model MKV1 defined a partition made up of 23 communities. Their average dimension is 23,43 with a relatively high standard deviation (16,63) due to the presence of three big communities (C0, C1 and C3) to which respectively 71, 57 and 47 nodes belong. All the agents of the initial set (352) are included in the considered partition and the majority of them (258) belong to just one community. Connections among communities are frequent, since the average number of overlaps per community is 6,47, and they are usually very small (average number of nodes per overlap equal to 2,10 with standard deviation equal to 1,42). Thus, the partition seems to reveal a framework characterized by a uniform distribution of weak connections. This is exactly what can be understood through the

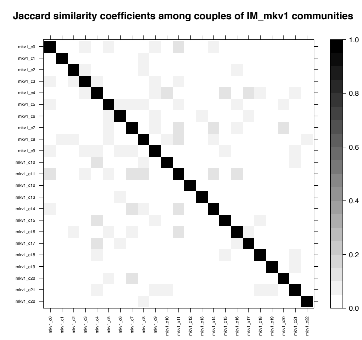



*Figure 6.11. Matrix of Jaccard similarity coefficients calculated for all pairs of model MKV1 communities. The darker is the color, the higher is the ratio. Rows and columns represent (ordered from the top to the bottom and from left to right) MKV1 communities.*

[91] In the case of model MKV2_AS, the condition that allows the circulation of the flow from agents that participated in a projects to agents that participated in another project is that *the beginning of the first project has to be anterior to the end of the second project*. With this condition, every participants could give rise to a flow of information starting from the moment of beginning of the project it is involved in.

[92] In the case of model MKV2_DS, the condition that allows the circulation of the flow from agents that participated in a projects to agents that participated in another project is that *the end of the first project has to be anterior to the end of the second project*. With this condition, every participants could give rise to a flow of information starting from the end of the project it is involved in.

*Table 6.12 (below). Adjusted residuals of the Contingency table regarding MKV1 communities (rows) and policies waves (columns) in which were collocated the projects to which agents took part. Since communities can overlap, a same agent can belong to more than one community. Thus, in every community to which an agent belongs, all its participations are considered. In green, adjusted residuals higher than or equal to 7. On the right, counting of the cells (by row) with adjusted residual higher than or equal to 7.*

| Adj. Res. | Waves | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **MKV1 Comm.** | 2002_171 | 2002_172 | 2002_ITT | 2004_171 | 2004_171A | 2005_171 | 2006_VIN | 2007_171 | 2008_171 | **Number of cells with Adj. Res. ≥ 7 by row** |
| **C0** | *6,03* | *-0,99* | *-4,16* | *-2,78* | *-1,16* | *5,25* | *-1,16* | *-4,08* | *3,52* | |
| **C1** | *8,17* | *-0,54* | *-3,68* | *6,94* | *-2,91* | *-2,74* | *-1,84* | *0,75* | *-0,47* | 1 |
| **C2** | *-3,36* | *0,57* | *-4,48* | *1,53* | *3,63* | *5,21* | *-1,38* | *-2,54* | *-0,35* | |
| **C3** | *-2,44* | *-3,48* | *12,88* | *-1,20* | *-2,61* | *-8,09* | *0,89* | *-0,21* | *-0,44* | 1 |
| **C4** | *-3,21* | *4,88* | *-4,35* | *0,66* | *-2,55* | *3,40* | *0,31* | *2,14* | *-0,34* | |
| **C5** | *3,51* | *-3,28* | *-2,94* | *-1,49* | *-0,99* | *4,62* | *-1,38* | *-0,34* | *-0,35* | |
| **C6** | *-1,46* | *2,46* | *-4,30* | *1,25* | *-2,15* | *4,01* | *-1,11* | *1,16* | *-0,29* | |
| **C7** | *1,37* | *-2,33* | *-4,21* | *-2,13* | *2,69* | *4,70* | *-0,98* | *0,60* | *-0,25* | |
| **C8** | *5,60* | *1,33* | *-3,79* | *-2,53* | *9,52* | *-1,98* | *-1,16* | *-2,13* | *-0,30* | 1 |
| **C9** | *-2,50* | *1,69* | *-5,84* | *-0,26* | *-0,89* | *6,60* | *-1,03* | *1,57* | *-0,26* | |
| **C10** | *-0,73* | *1,19* | *-0,12* | *-2,25* | *-2,00* | *-0,22* | *8,11* | *0,39* | *-0,27* | 1 |
| **C11** | *-0,61* | *-2,39* | *-5,17* | *1,86* | *3,67* | *3,55* | *-1,00* | *2,27* | *-0,26* | |
| **C12** | *-1,58* | *-1,54* | *5,77* | *-1,41* | *0,45* | *-2,90* | *-0,65* | *-1,19* | *-0,17* | |
| **C13** | *-2,67* | *1,69* | *6,01* | *-2,39* | *-2,13* | *-3,58* | *-1,10* | *1,22* | *-0,28* | |
| **C14** | *-1,44* | *-1,40* | *-1,93* | *-1,28* | *-1,14* | *4,58* | *2,89* | *-1,08* | *-0,15* | |
| **C15** | *-2,25* | *-0,17* | *-2,93* | *0,72* | *-1,79* | *4,74* | *-0,92* | *0,84* | *-0,24* | |
| **C16** | *-2,81* | *-2,74* | *1,93* | *2,81* | *9,10* | *-5,88* | *3,43* | *0,98* | *-0,30* | 1 |
| **C17** | *-2,77* | *3,53* | *4,86* | *4,70* | *-2,21* | *-6,50* | *-1,14* | *-0,01* | *-0,29* | |
| **C18** | *-0,08* | *9,00* | *-3,42* | *-1,92* | *-1,71* | *0,05* | *-0,88* | *1,03* | *-0,23* | 1 |
| **C19** | *-1,74* | *-1,70* | *6,64* | *1,24* | *-1,38* | *-4,07* | *-0,71* | *-1,31* | *-0,18* | |
| **C20** | *-2,39* | *-2,33* | *7,38* | *-2,13* | *-1,90* | *-4,50* | *-0,98* | *4,20* | *-0,25* | 1 |
| **C21** | *-2,79* | *-1,49* | *9,76* | *-2,50* | *-2,22* | *-5,83* | *4,39* | *-1,07* | *-0,29* | 1 |
| **C22** | *-0,65* | *-0,64* | *-1,66* | *-0,58* | *9,67* | *-1,53* | *-0,27* | *-0,49* | *-0,07* | 1 |

*Table 6.13 (below). Adjusted residuals of the Contingency table regarding MKV1 communities (rows) and the technological domains (columns) of the projects to which agents took part. Since communities can overlap, a same agent can belong to more than one community. Thus, in every community to which an agent belongs, all its participations are considered. In green, adjusted residuals higher than or equal to 7. On the right, counting of the cells (by row) with adjusted residual higher than or equal to 7.*

| Adj. Res. | Technological Domains | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **MKV1 Comm.** | Others | Bio-Technologies | Organic Chemistry | Geotermal Sciences and Biomasses | ICT - Multimedia | Mechanics Engineering | Multi Disciplinary | Nanotechnologies | New Materials | Optoelectronic | NA | **Number of cells with Adj. Res. ≥ 7 by row** |
| **C0** | *-0,51* | *-7,24* | *-1,49* | *0,05* | *9,32* | *-2,01* | *-1,76* | *1,44* | *1,36* | *-4,00* | *0,20* | 1 |
| **C1** | *-1,50* | *-1,96* | *0,68* | *-3,28* | *2,58* | *13,37* | *2,37* | *-2,60* | *-0,95* | *-6,96* | *-2,13* | 1 |
| **C2** | *-1,12* | *-4,50* | *2,03* | *-0,64* | *0,16* | *3,49* | *-1,10* | *7,59* | *-1,54* | *-0,98* | *-1,60* | 1 |
| **C3** | *-1,39* | *0,77* | *-2,12* | *-1,53* | *-11,25* | *-3,85* | *-1,36* | *0,39* | *-1,90* | *19,28* | *0,28* | 1 |
| **C4** | *-1,08* | *-2,61* | *-1,84* | *1,89* | *3,90* | *-1,45* | *-1,05* | *-1,86* | *-1,47* | *0,67* | *-1,53* | |
| **C5** | *-1,12* | *2,30* | *-2,80* | *-0,64* | *5,55* | *-0,91* | *-1,10* | *0,30* | *-1,54* | *-5,49* | *-0,24* | |
| **C6** | *-0,91* | *-0,99* | *-2,26* | *2,42* | *1,59* | *5,51* | *-0,88* | *-1,57* | *-1,24* | *-3,85* | *2,01* | |
| **C7** | *-0,80* | *-3,20* | *-1,99* | *3,21* | *2,50* | *0,80* | *-0,78* | *-1,38* | *-1,09* | *-0,98* | *2,57* | |
| **C8** | *-0,95* | *-0,61* | *0,47* | *-2,06* | *0,43* | *2,10* | *-0,92* | *4,26* | *9,41* | *-4,62* | *-1,34* | 1 |
| **C9** | *-0,84* | *-2,28* | *4,77* | *-1,82* | *2,01* | *-2,31* | *-0,81* | *-1,45* | *-1,14* | *-1,60* | *5,92* | |
| **C10** | *-0,84* | *-2,31* | *-2,09* | *-1,84* | *6,48* | *-0,90* | *-0,82* | *-1,46* | *-1,15* | *-1,65* | *-1,20* | |
| **C11** | *-0,82* | *-3,27* | *9,26* | *-1,78* | *-1,36* | *-1,28* | *3,13* | *0,08* | *12,07* | *-3,36* | *-1,16* | 2 |
| **C12** | *-0,53* | *9,47* | *0,32* | *0,68* | *-3,23* | *-1,46* | *-0,52* | *-0,92* | *-0,72* | *-2,58* | *-0,75* | 1 |
| **C13** | *20,22* | *-3,58* | *-2,23* | *-1,95* | *-6,88* | *-2,48* | *-0,87* | *-0,17* | *-1,22* | *11,96* | *-1,27* | 2 |
| **C14** | *-0,48* | *-1,92* | *-1,20* | *-1,05* | *4,39* | *-1,33* | *6,07* | *-0,83* | *-0,66* | *-2,35* | *-0,68* | |
| **C15** | *-0,75* | *0,52* | *-1,87* | *1,63* | *-1,92* | *-2,08* | *10,58* | *-1,31* | *-1,03* | *2,49* | *-1,07* | 1 |
| **C16** | *-0,94* | *7,10* | *9,01* | *8,05* | *-7,34* | *-1,31* | *-0,92* | *1,01* | *-1,28* | *-4,59* | *-1,34* | 3 |
| **C17** | *-0,93* | *8,56* | *-2,31* | *6,58* | *-4,13* | *-2,57* | *-0,91* | *-1,61* | *-1,27* | *-0,59* | *-1,32* | 1 |
| **C18** | *-0,72* | *5,33* | *-1,79* | *-1,57* | *0,85* | *-1,99* | *-0,70* | *-1,25* | *-0,98* | *-1,01* | *-1,02* | |
| **C19** | *-0,58* | *-2,33* | *1,53* | *-1,27* | *4,41* | *-1,61* | *-0,57* | *2,10* | *-0,79* | *-2,84* | *-0,83* | |
| **C20** | *-0,80* | *15,76* | *-1,99* | *-1,74* | *-6,14* | *-2,21* | *-0,78* | *-1,38* | *-1,09* | *-3,90* | *8,13* | 2 |
| **C21** | *-0,93* | *-1,82* | *-2,32* | *-2,04* | *-4,42* | *-2,59* | *-0,91* | *-1,62* | *-1,28* | *12,81* | *0,27* | 1 |
| **C22** | *-0,22* | *-0,87* | *9,23* | *-0,48* | *-1,96* | *-0,60* | *-0,21* | *-0,38* | *-0,30* | *-1,07* | *-0,31* | 1 |

observation of the graphic representation of the matrix in which the Jaccard similarity coefficients for all pairs of MKV1 communities have been calculated (figure 6.11). The matrix is scattered and the ratios between the intersection and the union set of all possible couples of communities do not reach interesting values.
The most interesting aspects regarding this model, as all other Infomap models, concern the observation of the contingency tables, especially (i) those regarding agents' participations in the different waves, and (ii) those regarding agents' participations in projects with specific technological domains. As can immediately be observed in table6.12 and in table 6.13, the communities detected by model MKV1 have particular characterizations both in terms of the waves in which agents participated and in terms of the technological domains of the projects in which agents participated. To have a clearer vision of these characterizations, cells with an adjusted residual higher than or equal to 7 have been colored in green[93].
As it is possible to observe, the characterization in terms of the waves appears to have a wide spread. 9 communities out of 23 show an adjusted residual higher than 7, and there are 4 waves to which there is not a specific correspondence with any community. On the other hand, 14 communities (out of a total of 23) show a high characterization in terms of at least one technological domain and, moreover, at least one community with that kind of specialization corresponds to every technological domain. Thus, this first model seems to be well oriented to highlight communities that have an interesting characterization from the perspective of the technological domains of the projects in which their agents participated. The following models will provide other elements to elaborate a more complete interpretation of what was observed in this partition.

### 6.2.2. MKV2_AS

Moving from model MKV1 to model MKV2_AS, the first element that immediately emerges is an explosion in the number of communities. While in the previous model 23 communities were found, here 54 communities are detected. As can be noticed in the right tail of the chart 6.14, half of them have less than 10 nodes, so for the most part they are communities with small dimensions. An average dimension of 14,13 and a standard deviation of 12,73 testify to the detection of a partition characterized by few communities that include a high number of nodes and a large quantity of communities with few nodes. From the point of view of the connections among communities, the situation does not change with regard to the previous model. There is a large number of overlaps (9,11 overlaps on average per community) and they all involve few nodes (average number of nodes per overlap equal to 1,81 and standard deviation equal to 1,43). Thus, the overlaps are scattered and a serried but light connective pattern emerges (figure 6.15).
The most interesting results of model MKV2_AS regard the contingency tables of agents' participations in different waves and of agents' participations in projects with specific technological domains (table 6.16). Even if with respect to what was found in model MKV1 the number of communities has more than doubled, the number of communities that are

[93] In the contingency tables that were presented before, adjusted residuals higher than or equal to 5 were highlighted. Here, the choice to stress more intensively the conditional formatting of the adjusted residuals deals with the necessity to focus only on the highest values. Lower thresholds would have caused a more difficult interpretation of the tables.

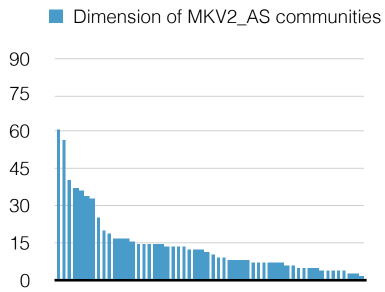



Chart 6.14 (left). Histogram representing the dimension (number of agents included) of MKV2_AS communities. The communities are in the same order that they have in the figure below the chart.

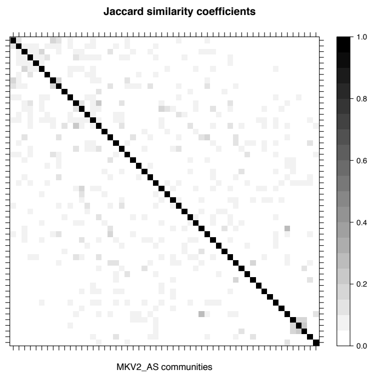



Figure 6.15 (left). Matrix of Jaccard similarity coefficients calculated for all pairs of model MKV2_AS communities. The darker is the color, the higher is the ratio. Rows and columns represent (ordered from the top to the bottom and from left to right) CPM_k05 communities. The communities are in the same order that they have in the chart above the figure.

| Number of communities with at least one adjusted residuals higher than or equal to 7 in the context of the corresponding contingency table | MKV2_AS | % out of the total number of communities |
|---|---|---|
| Typologies (agents) | 1 | *1,85%* |
| Ateco 2002 Sottosezioni (agents) | 1 | *1,85%* |
| Dimensional Classes (agents) | 0 | *0,00%* |
| Waves (participations) | 25 | *46,30%* |
| Technological domains (participations) | 26 | *48,15%* |
| **Number of communities in the model** | **54** | |

*Table 6.16 (above). Counting and percentages (over the total number of communities detected in the model) of MKV2_AS communities with at least one combinations (in the contingency table with the corresponding category indicated in the rows) characterized by an adjusted residual higher than or equal to 7. In the first column, in brackets, is indicated if the category taken into account refers to attributes of the agents or of their participations. In the last row, the total number of communities detected.*

characterized by significative levels of adjusted residuals (in the two just mentioned contingency tables) is high. In both cases, almost 50% of communities have an adjusted residual higher than or equal to 7. This evidence has to be observed with interest since, even if the number of communities is very large, a large part of them maintain strong and specific characterizations under the two considered profiles.

### 6.2.3. MKV2_DS

MKV2_DS, the Infomap model with the most restrictive limits to the circulation of the simulated flow, produced a partition that can be considered aligned with the trend depicted by the two preceding models. 62 communities were detected and they show features that are very similar to those of the partition identified by the model MKV2_AS. The average dimension of the communities is 8,21, with a standard deviation that is equal to 12,37. 50% of communities have fewer than 10 nodes and there are no longer isolated communities whose dimensions can be regarded as a peak (in the partition generated by MKV2_AS,

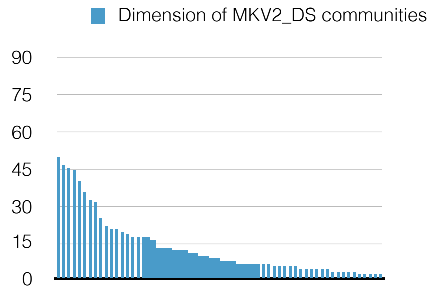



*Chart 6.17 (left). Histogram representing the dimension (number of agents included) of MKV2_DS communities. The communities are in the same order that they have in the figure below the chart.*

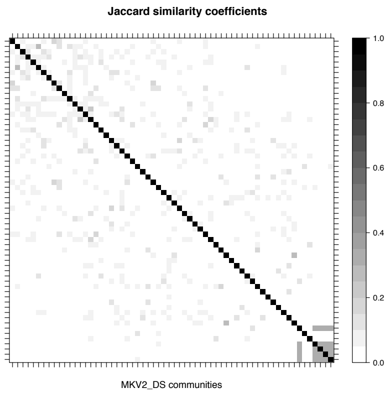



*Figure 6.18 (left). Matrix of Jaccard similarity coefficients calculated for all pairs of model MKV2_DS communities. The darker is the color, the higher is the ratio. Rows and columns represent (ordered from the top to the bottom and from left to right) CPM_k05 communities. The communities are in the same order that they have in the chart above the figure.*

| Number of communities with at least one adjusted residuals higher than or equal to 7 in the context of the corresponding contingency table | MKV2_DS | % out of the total number of communities |
|---|---|---|
| Typologies (agents) | 1 | *1,61%* |
| Ateco 2002 Sottosezioni (agents) | 2 | *3,23%* |
| Dimensional Classes (agents) | 0 | *0,00%* |
| Waves (participations) | 25 | *40,32%* |
| Technological domains (participations) | 31 | *50,00%* |
| **Number of communities in the model** | **62** | |

*Table 6.19 (above). Counting and percentages (over the total number of communities detected in the model) of MKV2_DS communities with at least one combinations (in the contingency table with the corresponding category indicated in the rows) characterized by an adjusted residual higher than or equal to 7. In the first column, in bracket, is indicated if the category taken into account refers to attributes of the agents or of their participations. In the last row, the total number of communities detected in the corresponding model.*

communities C0 and C1 have almost 60 agents, as can be noticed in chart 6.14). The average number of connections per community is 8,21 and the overlaps are made up of an average number of nodes equal to 1,76 (standard deviation equal to 1,38). These aspects find an adequate graphic representation in the matrix of the Jaccard similarity coefficients (figure 6.18) that, as in other Infomap models, are scattered and do not present specific elements of interest. The only peculiarity lies in a little group of communities that are intensively connected among them (right bottom corner of figure 6.18) but they all include such a small number of nodes (4 communities made up of 2 nodes and 1 community made up of 3 nodes) that they cannot be regarded with interest.

Finally, once again as for other Infomap models, the observation of the adjusted residuals reveals an intense characterization of these communities in terms of agents' participations in different waves and in terms of agents' participations in projects with specific technological domains. 50% of communities continue to have at least one adjusted residual higher than 7 in the contingency table that regards the technological domains of projects (table 6.19). Also the characterization in terms of participations in different waves is solid, but smaller than in model MKV2_AS. In model MKV2_DS the communities with this kind of characterization (at least one adjusted residual higher than or equal to 7 in the contingency table regarding technological domains of projects) are 40,32% (table 6.19) of the total number of communities detected, while in model MKV2_AS they are 46,30% (table 6.16).

### 6.2.4.Comparison of the three Infomap models

The most interesting elements about Infomap models come from the comparison of them. A kind of continuity over the three models seems to be present, especially for the fact that the more intensive the applied restriction was, the larger the number of communities is which present strong characterizations (thanks to the elaboration of the described contingency tables that I carried out). Also from the perspective of the communities detected, moving from MKV1 to MKV2_AS and then from MKV2_AS to MKV2_DS, an increase in their number was observed. This aspect is not only an element of strong interest but also an element of difficulty. In fact, the interpretation of partitions made up of a large number of communities prevented a clear vision of these partitions and, moreover, the patterns of connections are always serried, homogeneous but light. However, the presence of such a strong characterization in terms of agents' participations in projects with specific technological domains seems to suggest, step by step, the presence of a relation between Infomap methodology and the projects that were funded during the policies.

A better understanding of the results of Infomap models has to start necessarily from the comparison among the communities that were detected in the different models. To do this, I computed the similarity Jaccard coefficient for all communities belonging to any of the Infomap partitions (figure 6.20). The results deserve attention. Having highlighted three different colored areas which respectively represent (i) the ratios between MKV1 communities and MKV2_AS communities (yellow rectangle), (ii) the ratios between MKV1 communities and MKV2_DS communities (orange rectangle), and (iii) the ratios between MKV2_AS communities and MKV2_DS communities (dark orange rectangle), a peculiarity emerges in all these three areas. In every one of them, a series of cells with high ratios (cells colored in dark grey/black) surrounded by few other relevant combinations finds an oblique collocation that makes it possible to understand how all the communities detected in the model with less restrictive hypotheses have a specific correspondence with one community (or only few of them) detected by the model with more intense restrictions. In other words, in every highlighted area, every column has at least one cell where the ratio is close to 1. This is more evident between MKV2_AS communities and MKV2_DS communities (dark orange area) than in other areas. Probably because of the fact that MKV1 did not impose any limit on the circulation of the flow while the other two models did, the correspondence that MKV1 communities find with those belonging to MKV2_AS and MKV2_DS is more fuzzy, In addition, moving from MKV1 to MKV2_AS and then from MKV2_AS to MKV2_DS, a peculiar increasing similarity emerges along with an increase in the number of communities detected. This trend seems to depict a process of focusing into something that is still not clear. However, some evidence was present from the beginning.

High levels of characterizations of the communities in terms of agents' participations in different waves and in projects with different technological domains were observed starting

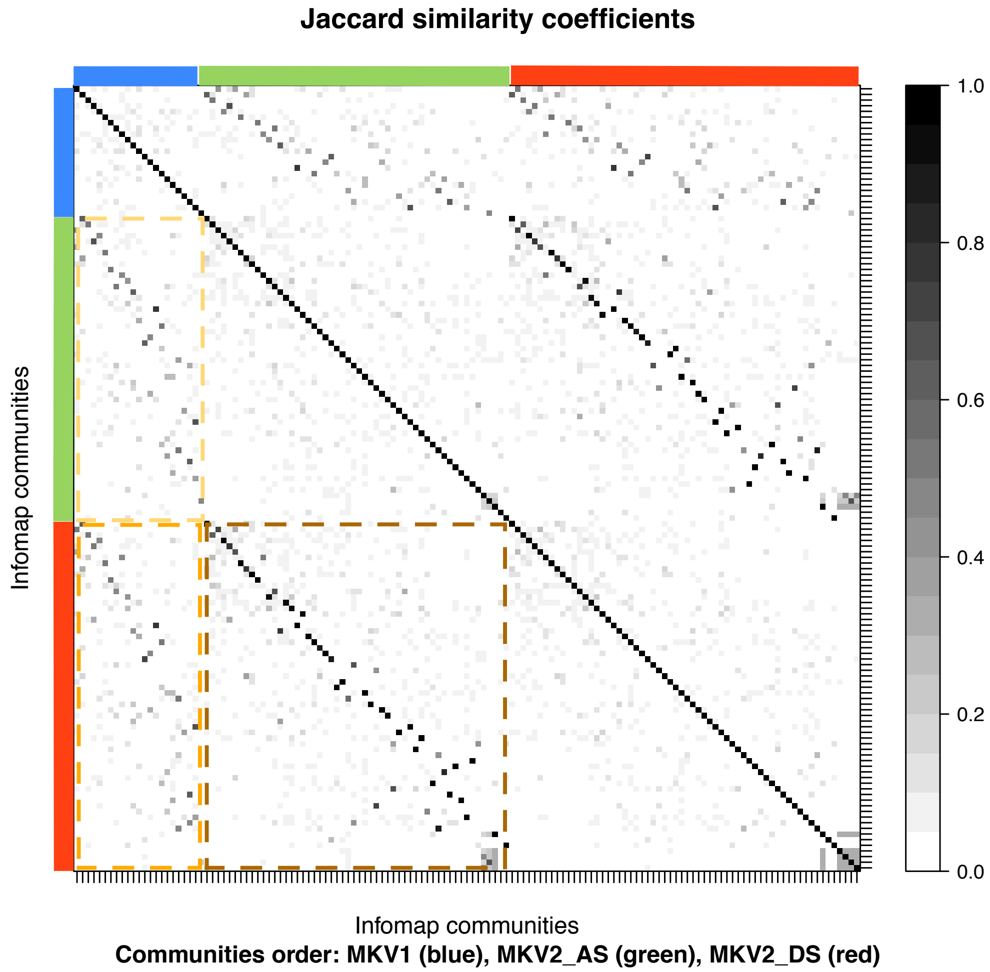



*Figure 6.20. Matrix of Jaccard similarity coefficients calculated for all pairs of communities detected in all three Infomap models. The darker is the color in the cell, the higher is the ratio. Yellow dashed rectangle highlights combinations between MKV1 and MKV2_AS communities. Orange dashed rectangle highlights combinations between MKV1 and MKV2_DS communities. Dark orange dashed rectangle highlights combinations between MKV2_AS and MKV2_DS communities.*

from model MKV1, and in following models they became stronger. These two different characterizations have a relevant element in common: they both regard the participations in which agents took part. Nevertheless, they put in evidence different aspects. Considering waves, the characterization seems to have to do with a specific collocation of the activity in time, or with some specific features/constraints of the waves themselves. On the other hand, the consideration of a characterization concerning technological domains seems to have more to do with the specific projects that were realized. In this context, an important element has to be remembered: while there were 9 waves, there were 168 projects. Thus, the observation of high levels of characterization regarding the technological domains in such large numbers of communities (MKV1 detected 23 communities, MKV2_AS 54 and MKV2_DS 62) seems to suggest that these communities are more oriented to reveal a partition devoted to giving rise to the projects' dynamics. For sure, this consideration is in

line with the functioning of the methodology, but there were no guarantees that this aspect would emerged.
To have a final confirmation of these elements, I computed the similarity Jaccard coefficient between the communities detected (in all Infomap models) and the partnerships of the 168 projects. Results are presented in figure 6.21. It can be observed that moving from MKV1 to MKV2_AS and then from MKV2_AS to MKV2_DS more intense similarities are observed. Model MKV2_DS (red circles) is the one that has the largest number of communities which perfectly overlap with one of the projects' partnerships. The fact that the model with the more restrictive conditions is the one that more than the others is able to investigate the processes (projects) that were realized in the context of the public policies, seems to be evidence of the coherence between the methodology and the aim to which it was applied. Infomap, the methodology that works on the processes that occurred (or could have occurred) in the network, led to the identification of communities that have a strong correspondence with the partnerships of the projects that were realized. This is interesting since it can be claimed that the processes that truly occurred in the network were the projects but, of course, this result could not be predicted before the analysis was made.
To conclude, one final consideration. Also CPM models show an interesting level of overlaps with the observed partnerships (figure 6.10). This fact seems to be obvious since partnerships generated connections. However, while CPM_k18, the CPM model that has to be regarded with most interest, has shown a correspondence with the waves also because of

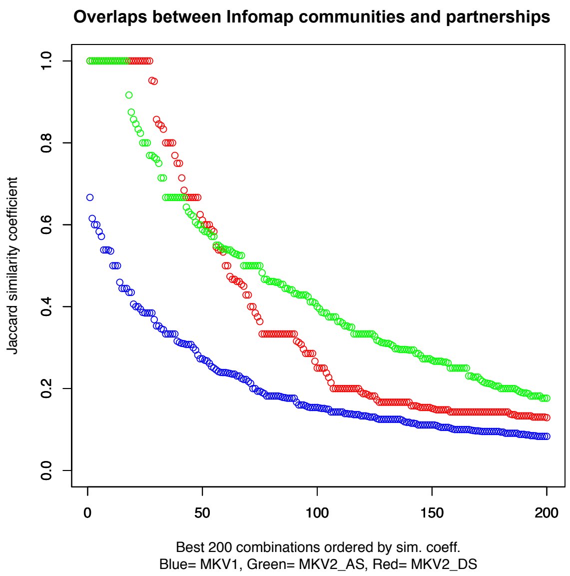



*Figure 6.21. Scatterplot of the Jaccard similarity coefficients calculated between couple made up by (i) one of the communities detected in one of the Infomap models (distinguished by different colors), and the partnerships configurations of the 168 projects that were realized in the context of the public policies. A coefficient equal to 1 indicates the perfect correspondence between the agents that belong to one of the communities detected and the agents involved in one of the observed projects. The calculated ratios were sorted in decreasing order and the best 200 coefficients were plotted. Every series has one circle for every position of the ranking (that is the X-axis). Thus, even if because of graphical computational limits does not appear so, the red circles with ratio equal to 1 do not start when the green ones finish but they start in correspondence of value 1 of the X-axis.*

the number of communities it detected (9 waves in the public policies and 10 communities in CPM_k18 model), in all Infomap models the large number of communities produced makes it more difficult to find precise correspondences with the waves. This fact seems to suggest that, even if both worked on the network framework, the two methodologies were able to investigate two different aspects of it. On one hand, CPM methodology investigated the structure and produced communities characterized by agent's participations in specific waves. On the other hand, Infomap investigated processes and produced communities characterized by agents' participations in specific technological domains, that for sure have more to do with the projects than with the waves. Results are coherent with the purposes for which the methodologies were applied.

## 6.3. DCI analysis communities

In the context of DCI models there are considerations that come before the discussion of the communities detected. In fact, the first results that these models produced regard the several rounds that were computed using DCI analysis. To run the algorithm more times was considered necessary to realize a process of skimming of the initial set. Without it, the detection of meaningful communities would have been affected by difficulties, since results would have been much more confused. In fact, as happened in model BOOL_1[94], also the initial rounds of model BOOL_2 and of model ∂LoA made communities emerge which from an economic point of view cannot be seen as interesting.
The need to restrict the analysis was solved through the reiteration of the application of the DCI algorithm to following sets from which, one after the other, the best group detected at the previous stage was excluded. While in model BOOL_1 this procedure continued until the exhaustion of agents, in model BOOL_2 and in model ∂LoA the skimming procedure was stopped after some rounds. There were two reasons for this change in the procedure:

- the continuous skimming of the best mask detected obviously leads to partitions where overlaps can occur only if data are rearranged, as in model BOOL_1;
- since DCI methodology investigates the presence of communities through the observation of the entire set in which agents are included, the progressive skimming produces a continuous alteration of the set under analysis, with consequences on the process of detection of the communities, which, moreover, with this alteration cannot be compared in terms of $t_{ci}$.

Thus, in the discussion of each model, how this procedure was handled is first of all described and, only afterwards, are considerations regarding the structure of the partition and the communities' characterizations provided.

[94] See paragraph 5.5.1.3

### 6.3.1. BOOL_1

Model BOOL_1 is the exploration of the application of DCI methodology to a context of analysis that for several reasons[95] has to be considered completely new. Thus, the importance of model BOOL_1 lies most in the fact that it made it possible to discover those problematic aspects whose presence was not known before the beginning of the analysis. Only thanks to this new level of awareness, was it possible to elaborate solutions that allowed the following two models to be more stable and coherent.

Starting from the initial set of those 352 agents with at least two participations in funded projects, the running of the algorithm revealed immediately a mask, which from an economic perspective was rather disappointing: 19 small local government institutions and 22 enterprises, most of which participated only twice in the public policies. After having excluded these agents, the algorithm was run once more but again the results were not very interesting: 8 enterprises and 8 KIBS, most of which with the minimum number of participations (table 6.24 and 6.25). However, it was clear that, since these agents had a very simple profile of activity due to their scarce participations, they were immediately grouped and detected by the algorithm. Given this context, the choice that was made was to continue the skimming of the best mask/masks.

The continuation of the procedure was not always constant. The changes that were made are described hereunder:

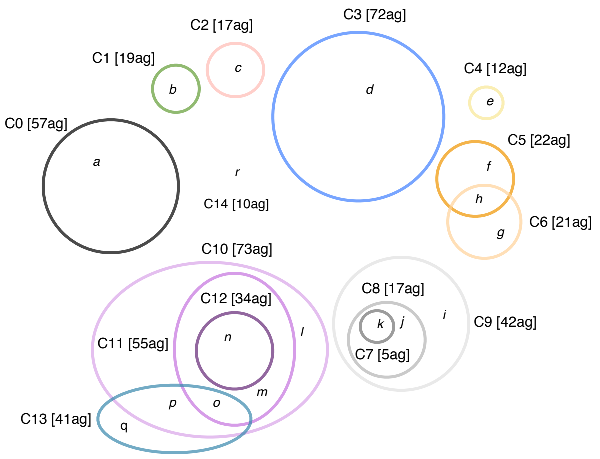
- first of all, the intent was to select the mask that, among all those detected in each run of the specific round, had the highest score in terms of $t_{ci}$. However, this criterion of selection had to adapt to the results that the algorithm produced. Sometimes it seemed to be more convenient to take only the best mask, sometimes the group of those agents that were always present in the masks with the highest scores, sometimes the group of those agents that were almost always present in the masks with the highest scores (depending on a given threshold);
- in round 6 and in round 7 more than one community was selected. The best masks of these rounds revealed the presence of two (in round 6) and of three (in round 7) groups that could well represent overlapping communities. For this reason more than one communities was cut out;
- the setting of the algorithm was not always constant because of some logistic and computational problems: the number of runs that were executed in every round was not always the same (at the beginning of the analysis only one run at a time could be launched, and the computational time of one of them took around five days) and the maximum subset dimension that was investigated by the algorithm had to be reduced after some rounds because of a too long a computational time;
- while in all rounds from the first to the eighth, the reiteration of the analysis was done over the initial set of the previous round minus the community (or communities) detected in the previous round (skimming procedure), in the ninth and in the tenth rounds the algorithm was launched on the communities that were detected respectively in the eighth and in the ninth rounds. The detection of two big communities both in round 8 (C10

[95] See paragraph 5.2.2

*Table 6.22 (below). Description of the settings (and of some other aspects) that characterized round by round the development of the BOOL_1 model. Notation 1: when the letter C, that stands for Community, is followed by square brackets, this represents a group of communities. For example, "C[0:3]" indicates all the communities which have name included between 0 and 3 (inclusive). Thus the group will be made up by C0, C1, C2 and C3. Notation 2: the maximum dimension (of subsets) investigated by the algorithm in round 3 was 110 in two runs, 90 in four runs, 80 in one run, 70 in two runs, 50 in two runs and 10 in one run.*

| BOOL_1 procedure | Rounds | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| Number of runs | 1 | 1 | 12 | 20 | 12 | 8 | 16 | 14 | 12 | 12 | 12 |
| Dimension of the initial set | 352 | 295 | 276 | 259 | 187 | 175 | 134 | 92 | 73 | 55 | 58 |
| Description of the agents that are present in the initial set | 352 agents with at least two participations in funded projects | 352 agents minus C0 | 352 agents minus C[0:1] | 352 agents minus C[0:2] | 352 agents minus C[0:3] | 352 agents minus C[0:4] | 352 agents minus C[0:6] | 352 agents minus C[0:9] | C10 | C11 | 352 agents minus C[0:9] minus C[12] |
| Maximum dimension (of subsets) investigated by the algorithm | 110 | 110 | 10 - 90 | 90 | 90 | 90 | 90 | 90 | 60 | 50 | 45 |
| Agetns selected at the end of the round | 57 | 19 | 17 | 72 | 12 | 41 | 42 | 73 | 55 | 34 | 41 |
| Number of detected subsystems | 1 | 1 | 1 | 1 | 1 | 2 | 3 | 1 | 1 | 1 | 1 |
| Dimension of the first detected subsystem | 57 | 19 | 17 | 72 | 12 | 22 | 5 | 73 | 55 | 34 | 41 |
| Dimension of the second detected subsystem | | | | | | 21 | 17 | | | | |
| Dimension of the third detected subsystem | | | | | | | 42 | | | | |
| Name of the detected communities | CO | C1 | C2 | C3 | C4 | C5 & C6 | C7 & C8 & C9 | C10 | C11 | C12 | C13 & C14 (residual agents) |

*Figure 6.23 (below). Graphical representation of the community detected in model BOOL_1 with the dimension (inside the square brackets) of each one of them. In the figure, letters distinguish different areas determined by the different overlaps (or not) of the communities. On the side, for each area is indicated the number of agents they are placed in it.*




| Areas in the representation | Number of agents |
|---|---|
| *a* | 57 |
| *b* | 19 |
| *c* | 17 |
| *d* | 72 |
| *e* | 21 |
| *f* | 20 |
| *g* | 19 |
| *h* | 2 |
| *i* | 25 |
| *j* | 12 |
| *k* | 5 |
| *l* | 4 |
| *m* | 3 |
| *n* | 34 |
| *o* | 18 |
| *p* | 14 |
| *q* | 9 |
| *r* | 10 |

| Obs. Freq. | Typology | | | | | | | | | Number of funded projects | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **BOOL_1 Comm.** | **GI** | **TA** | **CC** | **SC** | **LGI** | **E** | **R** | **BS** | **KIBS** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **17** | **18** | **TOT.** |
| **C0** | 1 | 3 | 7 | | 19 | 22 | | | 5 | 38 | 12 | 2 | | 2 | 2 | 1 | | | | | | | **57** |
| **C1** | 2 | | | | | 8 | 1 | | 8 | 15 | 3 | | 1 | | | | | | | | | | **19** |
| **C2** | 2 | | | | 4 | 6 | | 1 | 4 | 6 | 8 | | 1 | 1 | 1 | | | | | | | | **17** |
| **C3** | 2 | 3 | 3 | 1 | 5 | 17 | 7 | | 34 | 30 | 16 | 5 | 3 | 3 | 2 | 7 | 1 | 2 | 1 | 1 | | 1 | **72** |
| **C4** | | 1 | | | | 8 | 1 | | 2 | 11 | | 1 | | | | | | | | | | | **12** |
| **C5** | 1 | | | | 1 | 11 | | 1 | 8 | 10 | 5 | 4 | 1 | 2 | | | | | | | | | **22** |
| **C6** | | 1 | | | 1 | 3 | 4 | | 12 | 10 | 4 | 5 | 2 | | | | | | | | | | **21** |
| **C7** | | | | | 2 | 2 | | | 1 | 4 | 1 | | | | | | | | | | | | **5** |
| **C8** | | 1 | | | 5 | 6 | | 1 | 4 | 10 | 4 | 1 | 2 | | | | | | | | | | **17** |
| **C9** | 1 | 2 | | 1 | 5 | 15 | 4 | 2 | 12 | 20 | 8 | 4 | 8 | 1 | 1 | | | | | | | | **42** |
| **C10** | 4 | | | | 7 | 23 | 7 | 2 | 30 | 25 | 17 | 16 | 2 | 5 | 3 | 3 | 1 | | | | 1 | | **73** |
| **C11** | 4 | | | | 5 | 12 | 4 | 2 | 28 | 15 | 15 | 12 | 1 | 4 | 3 | 3 | 1 | | | | 1 | | **55** |
| **C12** | 3 | | | | 3 | 7 | 1 | 2 | 18 | 9 | 11 | 8 | | 2 | 2 | 2 | | | | | | | **34** |
| **C13** | 1 | | | | 4 | 18 | 5 | 1 | 12 | 18 | 7 | 7 | 2 | 4 | 1 | 1 | 1 | | | | | | **41** |
| **C14** | | | | | | 3 | 3 | | 4 | | | | | | 3 | 3 | | 4 | | | | | **10** |

*Table 6.24 (left). Contingency table regarding communities and the typology of the agents. The cells contain the observed frequencies of the agents that belong to the corresponding BOOL_1 community (rows) and that have the corresponding typology (columns). Since communities can overlap, a same agent can belong to more than one community. Thus any agent is considered in every community to which it belongs. In green, the cells to which correspond an adjusted residual higher than or equal to 1,5. The variable 'Typology' is made up by ten different categories: Government Institutions (GI), Trade Associations (TA), Chambers of Commerce (CC), Service Centers (SC), Local Government Institutions (LGI), Enterprises (E), Universities and Public Research Centers (R), Private Research Enterprises (PRC), Business Services Enterprises (BS), Knowledge-Intensitive Business Services (KIBS).*

*Table 6.25 (right). Contingency table regarding communities and the number of funded projects in which agents participated. The cells contain the observed frequencies of the agents that belong to the corresponding BOOL_1 community (rows) and that participated in the corresponding quantity of funded projects (columns). Since communities can overlap, a same agent can belong to more than one community. Thus any agent is considered in every community to which it belongs.*

includes 73 agents) and in round 9 (C11 includes 55 agents) suggested a change of direction: from the exclusion of the detected groups to the focusing on them;

- finally, in the last round (the eleventh) the reintroduction of those agents that were excluded in the passage from the eighth to the ninth round and of those that were excluded in the passage from the ninth round to the tenth, was carried out. Thus, the initial set of the last round was composed of: (i) those agents that were analyzed in the tenth round but did not belong to the selected group (21 agents), (ii) those agents that were analyzed in the ninth round but did not belong to the selected group (18 agents), and (iii) those agents that were analyzed in the eighth round but did not belong to the selected group (19 agents). Thanks to this rearrangement, overlaps emerged.

This is what was done in the context of this explorative analysis that in the end produced a partition characterized mostly by disjoint communities, as is natural for it to have been, having sees how it was realized. From the point of view of the characterization of the communities, the reading of the contingency tables can suggest some interesting considerations that are perfectly in line with the results that, from the beginning, had been observed. In fact, observing the contingency table that concerns the typology of agents that are present in the communities of the model BOOL_1 (table 6.24), it emerges that 13 communities out of the 15 detected have at least one observed frequency to which corresponds an adjusted residual higher than or equal to 1,5. Even if the considered

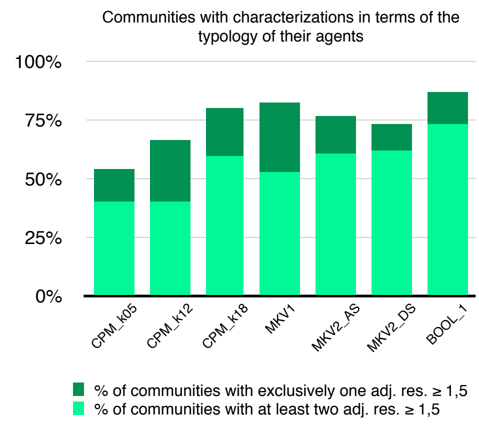



*Chart 6.26 (on the side above). Histogram of the percentages of the number of communities (over the total of communities detected in the corresponding model) which have exclusively one adjusted residual higher than or equal to 1,5 (light green) and percentages of the number of communities (over the total of communities detected in the corresponding model) which have at least two adjusted residual higher than, or equal to 1,5 (dark green), by partition to which they belong.*

| Characterizations in Typology of agents contingency tables | CPM_k05 | CPM_k12 | CPM_k18 | MKV1 | MKV2_AS | MKV2_DS | BOOL_1 |
|---|---|---|---|---|---|---|---|
| Number of communities with at least 2 adj. res. ≥ 1,5 | 2 | 7 | 2 | 7 | 8 | 7 | 2 |
| Number of communities with exclusively 1 adj. res. ≥ 1,5 | 6 | 11 | 6 | 12 | 33 | 38 | 11 |
| Number of communities | 15 | 27 | 10 | 23 | 54 | 62 | 15 |
| % of communities with at least 2 adj. res. ≥ 1,5 | 40,00% | 40,74% | 60,00% | 52,17% | 61,11% | 61,29% | 73,33% |
| % of communities with only 1 adj. res. ≥ 1,5 | 13,33% | 25,93% | 20,00% | 30,43% | 14,81% | 11,29% | 13,33% |

*Table 6.27 (on the side below). First row: number of communities, distinguished by the model in which they were detected (columns), with at least two adjusted residuals higher than or equal to 1,5. Second row: number of communities, distinguished by the model in which they were detected (columns), with exclusively one adjusted residuals higher than or equal to 1,5. Third row: total number of communities detected in the corresponding model (columns). Fourth row: ratio between the first row and the third. Fifth row: ratio between the second row and the third.*

threshold (adjusted residual higher than or equal 1,5) is not too high[96], especially if compared with the ones used before, it is important to underline that other methodologies show patterns of characterization that are less interesting under the profile regarding agent's typologies. In most of the cases, models produced (a) either a partition with a lower percentage of communities characterized by the typology of their agents, or (b) a partition to which belong communities with more than one characterization under this profile. As can be observed in chart 6.26, BOOL_1 has not only the highest percentage of communities with an adjusted residual higher than 1,5, but it has also the highest percentage of communities with exclusively one typology with an adjusted residual higher than 1,5. This suggests the presence (in BOOL_1 partition) of a pattern of specific characterizations regarding the typology of the agents.

To conclude, a final observation. The last community that was identified (C14) was not detected by the algorithm but it can be considered the final residual of it (these agents were grouped together because at the end of the analysis they were the only ones that did not belong to any community). It is interesting to notice that the 10 agents which belong to this group have an average of 9,20 participations in funded projects. The impression is that, because of their complex patterns of activity along with the limits of the data structure (with regard to the limited number of instants in time), the algorithm was not able to group them

[96] The setting of a higher threshold in the reading of adjusted residuals regarding the typologies of agents nullify almost all the characterizations in all the models.

together. However, they had such an intense participation in the considered public policies that, without any doubt, they deserved to take part in the partition.

### 6.3.2. BOOL_2

After having terminated BOOL_1 analysis, some important elements become absolutely of primary importance in order to define more stable and coherent criteria for the whole procedure that regards the application of DCI algorithm. First of all, the need to fix the number of the runs that have to be launched in every round was overcome thanks to an implementation of the software that allowed the automatic repetition of a series of runs and the computation of a single final ranking list (sorted by DCI significance). Thus, the number of runs per round was fixed at 30. Then, to avoid arbitrariness, it was decided to proceed by skimming in every round exclusively the best mask detected (the top one in the final ranking list), and the maximum dimension (of subsets) investigated was fixed at 90. Finally, the need to stop the progressive skimming procedure (that causes the incomparability of the masks in terms of $t_{ci}$ and that prevents the detection of overlapping communities) was solved thanks to the proposal I made about the introduction of a cluster analysis of the masks[97], a procedure that in such a context allows the identification of communities without proceeding with the alteration of the set under analysis (as happened in model BOOL_1). Moreover, the cluster analysis reveals two specific advantages: (i) first of all, communities can overlap among themselves; (ii) secondly, since masks are agglomerated on the basis of their similarities (simple matching), different clusters represent configurations of agents that are dissimilar among each other, and I retained this could be an appropriate way to investigate the presence of diversity among masks. Nevertheless, some initial skimmings were done. In fact, clearly BOOL_1 model showed that the first masks detected include agents that should have been removed before the beginning of the analysis. Since there is no way to do this without arbitrariness, skimmings were reiterated until the best detected mask involved agents that do not deserve to be excluded from the analysis.

Model BOOL_2 started with the consideration of those 352 agents with at least two participations in funded projects. The first three rounds of the algorithm revealed masks that do not seem to be interesting, and so they were skimmed one after the other, round by round. The fourth round highlighted a mask composed of 9 enterprises and 7 KIBS and because of the relevant typologies of the majority of its agents (that are in total 22) it was decided that it was not going to be skimmed (table 6.28). Thus, the 12.500 best masks (figure 6.29) detected by the algorithm at the end of the fourth round were clusterized (hierarchical agglomerative clustering) using the simple matching coefficient as similarity measure and the complete linkage as agglomerative criterion (figure 6.31). The observation of the differences of heights among the successive nodes in the dendrogram, and the consideration of the number of agents overall involved depending on the different cuts, suggested that I define a division into 13 clusters. In fact, the cut of the dendrogram in 13 clusters correspond to a local maximum in the differences of heights among successive nodes (figure 6.31) and it allows me to include 210 agents in the whole partition. For every one of these clusters - which are groups of masks and not groups of agents - the mask with the highest score in

[97] See paragraph 5.4

*BOOL_2: cluster analysis of masks*

| Typology | GI | TA | CC | SC | LGI | E | R | PRC | BS | KIBS | TOTAL (number of skimmed agents) | Agents remaining after the skimming of the best mask |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| round_1 | 1 | 2 | | | 18 | 16 | | | | 5 | **42** | *310* |
| round_2 | 2 | | | | | 8 | 1 | | | 8 | **19** | *291* |
| round_3 | 2 | | | | 3 | 6 | | | 1 | 4 | **16** | *275* |
| round_4 | | 1 | 1 | | 2 | 9 | 2 | | | 7 | **22** | |

*Table 6.28 (above). Agents that were detected in the best mask (the one with the highest tci) in the corresponding round of model BOOL_2 (on the rows). Agents are tabled on the basis of their typology. At the end of every round, the best mask was skimmed from the system and the analysis was reiterated. After the end of the seventh round, the skimming procedure was not made and so the agents which were present after the skimming of the best mask detected at the end of the sixth round (green cell) were considered for the rest of the analysis. The variable 'Typology' is made up by ten different categories: Government Institutions (GI), Trade Associations (TA), Chambers of Commerce (CC), Service Centers (SC), Local Government Institutions (LGI), Enterprises (E), Universities and Public Research Centers (R), Private Research Enterprises (PRC), Business Services Enterprises (BS), Knowledge-Intensitive Business Services (KIBS).*

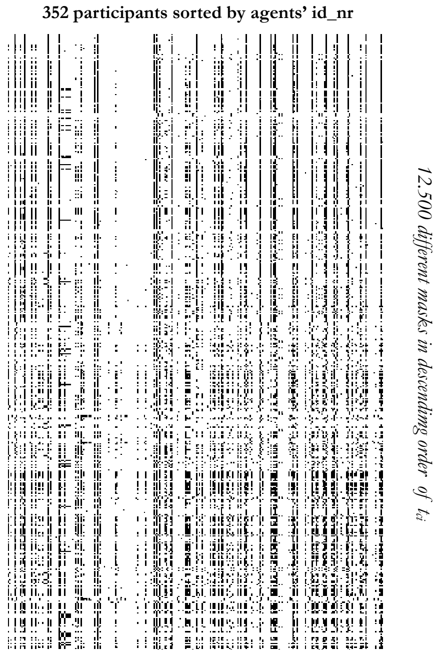
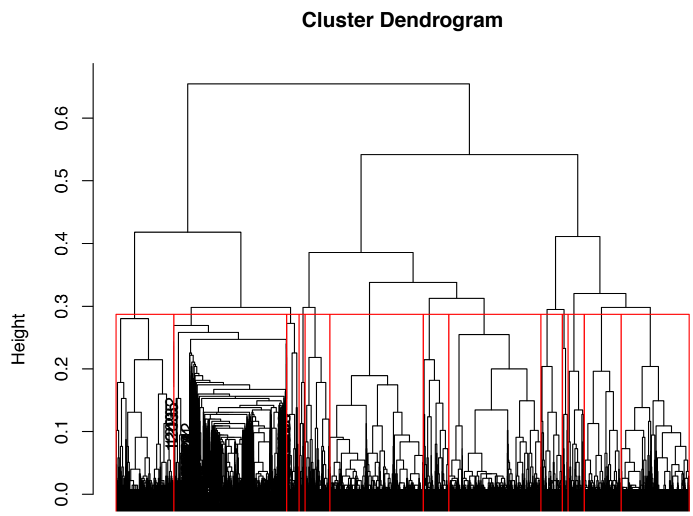
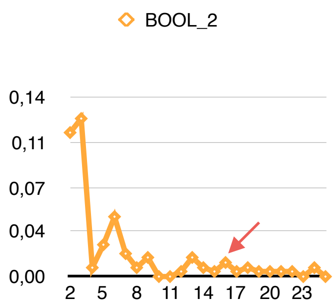



*Figure 6.29 (on the side). 12.500 best masks (subset of agents) detected at the end of the seventh round. Every row is a different masks and they are ordered in descending order of tci starting from the top. 275 agents were considered. Every column corresponds to an agent and when the cell is black it means that the corresponding agent (column) belongs to the corresponding subset (column).*

*Figure 6.30 (below). Dendrogram of the cluster analysis that took into account the best 12.500 masks detected at the end of the seventh round. Since a mask is a binary variable (equal to 1 if the corresponding agent belongs to the mask, equal to 0 if not) the distance matrix was computed using simple matching similarity coefficients. The complete linkage was used as criterion of agglomeration of the clusters. In the figure, the red rectangles illustrate the final clusters. It is important to remember that every cluster is a group of masks, and not a group of agents. After the cluster analysis, in every cluster the mask with the highest score in terms of tci was assumed as the representative mask of the group.*

*Chart 6.31 (below on the right). Differences of heights among successive nodes in the dendrogram. The numbers of the X-axis indicate the corresponding number of cluster. The red arrow shows where it was decided to cut the dendrogram. The 16 clusters determined the identification of 13 communities.*



| Number of agents | Typology | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| BOOL_2 Comm. | GI | TA | CC | SC | LGI | E | R | BS | KIBS | TOT. |
| C0 | | 1 | 1 | | 2 | 9 | 2 | | 7 | **22** |
| C1 | | 1 | | | | 7 | | | 3 | **11** |
| C2 | | | 7 | | 2 | 2 | 2 | | 8 | **21** |
| C3 | 1 | 4 | 4 | 1 | 5 | 22 | 6 | | 27 | **70** |
| C4 | | 2 | | | 2 | 6 | 1 | | 7 | **18** |
| C5 | 1 | 2 | | | 2 | 11 | 1 | | 4 | **21** |
| C6 | 2 | 3 | 4 | 1 | 4 | 23 | 7 | 1 | 26 | **71** |
| C7 | | 2 | | | 5 | 21 | 1 | 1 | 13 | **43** |
| C8 | 1 | | | | 2 | 3 | 2 | 1 | 5 | **14** |
| C9 | 2 | 4 | | 1 | 7 | 19 | 7 | | 29 | **69** |
| C10 | | 2 | 10 | | 5 | 23 | 12 | 1 | 29 | **82** |
| C11 | 2 | 1 | | 2 | 2 | 16 | 2 | | 9 | **34** |
| C12 | 2 | 3 | | 2 | 4 | 26 | 4 | 1 | 15 | **57** |

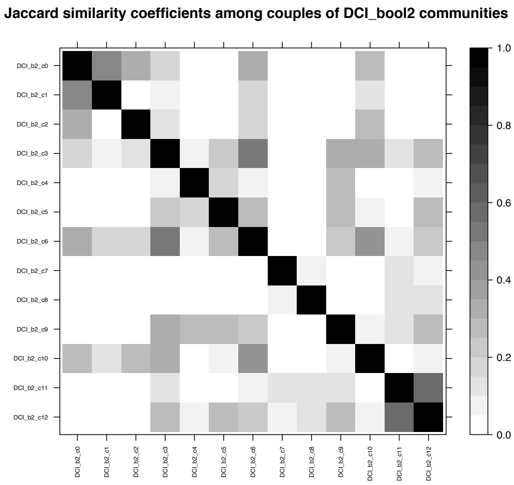



*Table 6.32 (left). Contingency table regarding communities and the typology of the agents. The cells contain the observed frequencies of the agents which belong to the corresponding BOOL_2 community (rows) and which have the corresponding typology (columns). Since communities can overlap, a same agent can belong to more than one community. Thus, any agent is considered in every community to which it belongs. The communities are in the same order that they have in the figure on the side. The variable 'Typology' is made up by ten different categories: Government Institutions (GI), Trade Associations (TA), Chambers of Commerce (CC), Service Centers (SC), Local Government Institutions (LGI), Enterprises (E), Universities and Public Research Centers (R), Private Research Enterprises (PRC), Business Services Enterprises (BS), Knowledge-Intensitive Business Services (KIBS).*

*Figure 6.33 (right). Matrix of Jaccard similarity coefficients calculated for all pairs of model BOOL_2 communities. The darker is the color, the higher is the ratio. Rows and columns represent (ordered from the top to the bottom and from left to right) BOOL_2 communities. The communities are in the same order that they have in the table on the side.*

terms of $t_{ci}$ was assumed as the representative of the whole cluster. In this way, 13 communities were identified.

The 13 communities detected have an average dimension of 41,00 agents and a standard deviation equal to 25,55. The pattern of intersections is very sparse but very intense. Communities have on average 3,69 connections with an average dimension of 14,13 agents (standard deviation equal to 11,87). The most interesting element is that the three biggest communities (C3, C6 and C10, which are respectively made up of 70, 71 and 82 nodes) are strongly overlapped among themselves (the three intersections have dimensions 50, 39 and 44) and they are all characterized by the presence of a large number of enterprises and KIBS (in table 6.32, the cells containing the total number of agents belonging to these communities are colored in orange). Among them, C10 (the biggest one) presents an element of peculiarity: the presence of 12 Universities/public research centers (C10 is the only community with a double-digit number of agents of this typology). The fourth biggest community (C9, which includes 69 agents and whose total number of agents is highlighted in yellow in table 6.32) seems to be more detached from the other biggest communities, since it has intense connections with C3 and C6 (respectively of size 34 and 26) but not with C10. Like the other biggest communities, C9 includes a large number of enterprises and KIBS (respectively 19 and 29). The remaining communities do not present aspects of particular interest and, also from the point of view of connections, they have a more marginal collocation.

From the observation of the contingency tables specific elements do not emerge but a final consideration has to be made. Even if the adjusted residuals that regard the typology of the agents show somewhat low values (only 8 communities out of 13, have at least one adjusted residual higher than 1,5), some elements were already detected. The fact that the four biggest communities are all characterized by a significant presence of enterprises and KIBS and, above all, that the biggest community is the only one with a double-digit number of Universities/public research centers should not pass unobserved.

### 6.3.3. ∂LoA model and considerations on DCI methodology

As in model BOOL_2, in model ∂LoA the same procedure that was developed after the preliminary and explorative model BOOL_1 was used. Starting from the set of agents that participated in at least two funded projects (352 agents), several rounds of the DCI algorithm were carried out. The skimming proceeded step by step with the exclusion of the best mask detected at the end of every round. At the end of the seventh round (table 6.34), even if the mask that was detected as the best is made up of only 4 agents, among them there is a KIBS that had 5 participations (out of 7 proposals) in funded projects and a local government institution that had 6 participations in funded projects (out of 9 proposals). Since these agents had intense activities, they deserved not to be excluded from the analysis and thus the skimming procedure was stopped here. The 259 agents which were present at the beginning of the seventh round were selected to constitute the set to analyze.
The ranking of the top 12.500 masks detected at the end of the seventh round (figure 6.35) was analyzed with a cluster analysis procedure (hierarchical agglomerative) that used a simple matching similarity measure to calculate the dissimilarity matrix and a complete linkage criterion to aggregate clusters step by step. The dendrogram was computed (figure 6.36) and, thanks to the observation of the differences of heights among the successive nodes in it (figure 6.37), and thanks to the consideration of the number of agents overall involved depending on the different cuts, I defined a division into 16 clusters. The cut of the dendrogram in 16 clusters corresponds to a local maximum in the differences of heights among successive nodes, and it allowed me to include 239 agents in the whole partition. For every one of them - which are groups of masks and not groups of agents - the mask with the highest score in terms of $t_{ci}$ was assumed as the representative of the whole cluster. In this way, 16 communities were identified.
The partition that model ∂LoA produced is characterized by four communities that reach the maximum dimension investigated by the algorithm (90 agents). On average, communities are made up of 50,06 agents (standard deviation equal to 33,91) and, without considering the first three of them (see C0, C1 and C2 in chart 6.38) which are very small, the minimum dimension observed is determined by 25 agents. Communities have on average 6,53 overlaps which on average contain 16,80 agents. The number of agents included in these intersections is not uniformly distributed and in fact its standard deviation, which is equal to 20,33, is superior to the average itself. This suggests the presence of a pattern of connections among communities that is not uniform and that becomes very dense in some particular combinations. This is exactly what can be observed in the graphic representation of the Jaccard similarity matrix that concerns the ratio between the intersection and the union set

| Typology | GI | TA | CC | SC | LGI | E | R | PRC | BS | KIBS | TOTAL (number of skimmed agents) | Agents remaining after the skimming of the best mask |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| round_1 | 1 | 2 | | | 18 | 16 | | | | 5 | 42 | *310* |
| round_2 | 2 | | | | | 8 | 1 | | | 8 | 19 | *291* |
| round_3 | 2 | | | | | 4 | 3 | | | 5 | 14 | *277* |
| round_4 | | 1 | | | | 4 | | | | 1 | 6 | *271* |
| round_5 | | 1 | | | | 5 | | | | 1 | 7 | *264* |
| round_6 | | | | | | | | | | 5 | 5 | *259* |
| round_7 | | | | | 2 | | | | 1 | 1 | 4 | |

*Table 6.34 (above). Agents that were detected in the best mask (the one with the highest $t_{ci}$) in the corresponding round of model ∂LoA (on the rows). Agents are tabled on the basis of their typology. At the end of every round, the best mask was skimmed from the system and the analysis was reiterated. After the end of the seventh round, the skimming procedure was not made and so the agents which were present after the skimming of the best mask detected at the end of the sixth round (green cell) were considered for the rest of the analysis. The variable 'Typology' is made up by ten different categories: Government Institutions (GI), Trade Associations (TA), Chambers of Commerce (CC), Service Centers (SC), Local Government Institutions (LGI), Enterprises (E), Universities and Public Research Centers (R), Private Research Enterprises (PRC), Business Services Enterprises (BS), Knowledge-Intensive Business Services (KIBS).*

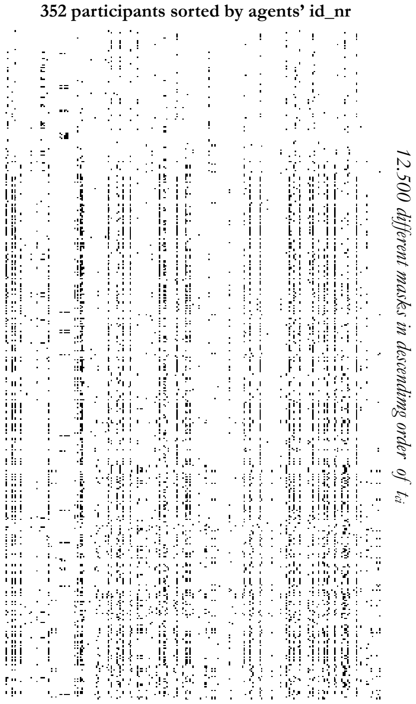
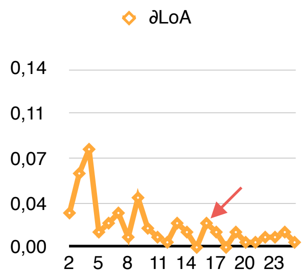
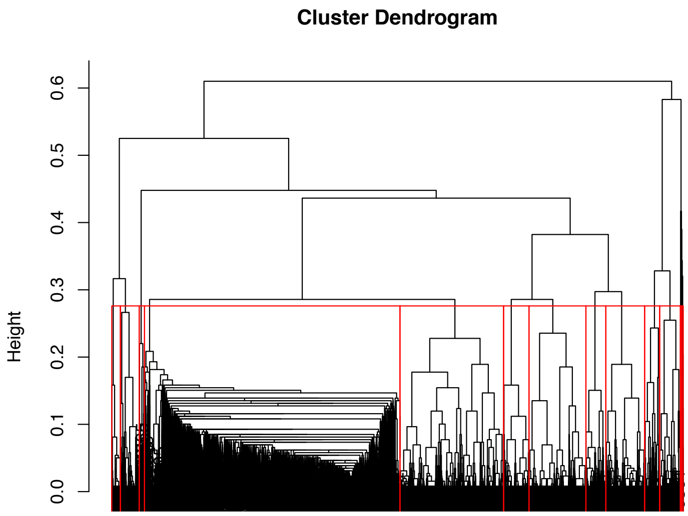



*Figure 6.35 (on the side). 12.500 best masks (subset of agents) detected at the end of the seventh round. Every row is a different masks and they are ordered in descending order of $t_{ci}$ starting from the top. 259 agents were considered. Every column corresponds to an agent and when the cell is black it means that the corresponding agent (column) belongs to the corresponding subset (column).*

*Figure 6.36 (below). Dendrogram of the cluster analysis that took into account the best 12.500 masks detected at the end of the seventh round. Since a mask is a binary variable (equal to 1 if the corresponding agent belongs to the mask, equal to 0 if not) the distance matrix was computed using simple matching similarity coefficients. The complete linkage was used as criterion of agglomeration of the clusters. In the figure, the red rectangles illustrate the final clusters. It is important to remember that every cluster is a group of masks, and not a group of agents. After the cluster analysis, in every cluster the mask with the highest score in terms of $t_{ci}$ was assumed as the representative mask of the group.*

*Chart 6.37 (below on the right). Differences of heights among successive nodes in the dendrogram. The numbers of the X-axis indicate the corresponding number of cluster. The red arrow shows where it was decided to cut the dendrogram. The 16 clusters determined the identification of 16 communities.*



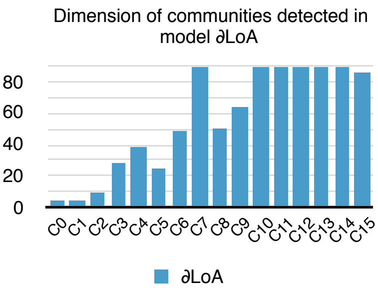



*Chart 6.38. Histogram representing the dimension (number of agents included) of ∂LoA communities.*

among all possible pairs of communities belonging to the model ∂LoA (figure 6.40). It is interesting to observe how two distinct areas can be clearly identified, and in fact in the graphic representation two squares are clearly discernible. The first, lighter than the second and located in the middle of the matrix, includes communities C3, C4, C6, C7, C8, C9 and C10. The second, very marked and located at the bottom right corner of the matrix, regards communities C11, C12, C13, C14 and C15. The interesting peculiarity is that these two groups of communities, besides being characterized by intense overlaps within their communities, are strongly characterized in terms of the typology of their agents. As can be observed in the table beside the matrix (yellow cells in table 6.39), all the communities that are located in the first described group have a large number of enterprises and KIBS (the average number of enterprises is 18,00 and the average number of KIBS is 17,63). On the other hand, the communities which belong to the second group are also characterized by the strong presence of enterprises and KIBS (the average number of enterprises is 19,40 and the average

| Number of agents | Typology | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **∂LoA Comm.** | **GI** | **TA** | **CC** | **SC** | **LGI** | **E** | **R** | **BS** | **KIBS** | **TOT.** |
| **C0** | | | | | 2 | | | 1 | 1 | **4** |
| **C1** | | 1 | | | 2 | | | 1 | 1 | **5** |
| **C2** | | | | | 2 | 5 | | 1 | 2 | **10** |
| **C3** | 2 | | | 1 | 2 | 12 | 3 | 1 | 8 | **29** |
| **C4** | 2 | 2 | | 1 | 4 | 15 | 4 | | 11 | **39** |
| **C5** | 1 | | 6 | | 2 | 4 | 3 | 1 | 8 | **25** |
| **C6** | 1 | 4 | | | 8 | 17 | 5 | 1 | 12 | **48** |
| **C7** | 2 | 3 | 2 | 1 | 11 | 25 | 8 | 2 | 35 | **89** |
| **C8** | 2 | 1 | | | 9 | 24 | 2 | 1 | 11 | **50** |
| **C9** | 3 | 2 | | 1 | 6 | 25 | 7 | 1 | 18 | **63** |
| **C10** | 2 | 1 | 8 | | 9 | 22 | 8 | 2 | 38 | **90** |
| **C11** | 2 | | 2 | 1 | 6 | 20 | 14 | 5 | 40 | **90** |
| **C12** | 2 | | 2 | 1 | 10 | 19 | 13 | 1 | 42 | **90** |
| **C13** | 1 | | 5 | 1 | 4 | 23 | 14 | 2 | 39 | **89** |
| **C14** | 3 | | 5 | 1 | 5 | 19 | 15 | 1 | 41 | **90** |
| **C15** | 2 | | 1 | 1 | 9 | 16 | 13 | 1 | 43 | **86** |

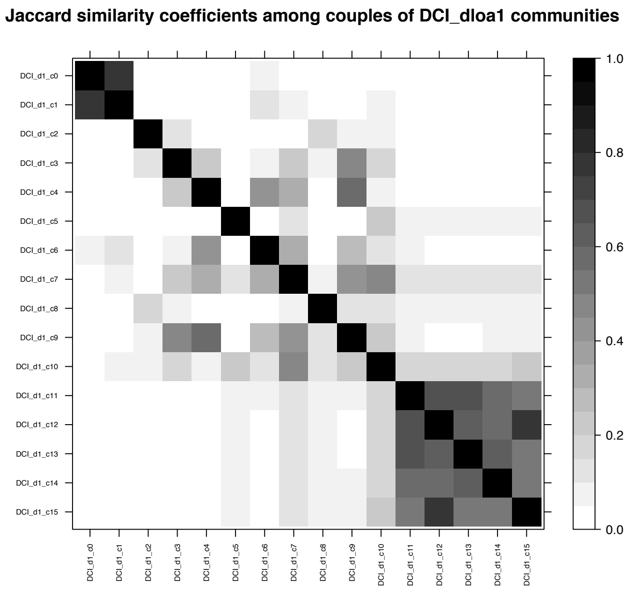



*Table 6.39 (left). Contingency table regarding communities and the typology of the agents. The cells contain the observed frequencies of the agents which belong to the corresponding ∂LoA community (rows) and which have the corresponding typology (columns). Since communities can overlap, a same agent can belong to more than one community. Thus, any agent is considered in every community to which it belongs. The communities are in the same order that they have in the figure on the side. The variable 'Typology' is made up by ten different categories: Government Institutions (GI), Trade Associations (TA), Chambers of Commerce (CC), Service Centers (SC), Local Government Institutions (LGI), Enterprises (E), Universities and Public Research Centers (R), Private Research Enterprises (PRC), Business Services Enterprises (BS), Knowledge-Intensitive Business Services (KIBS).*

*Figure 6.40 (right). Matrix of Jaccard similarity coefficients calculated for all pairs of model ∂LoA communities. The darker is the color, the higher is the ratio. Rows and columns represent (ordered from the top to the bottom and from left to right) ∂LoA communities. The communities are in the same order that they have in the table on the side.*

number of KIBS is 41,00) but clearly another typology of agents emerges: Universities and public research centers ('R' category). The orange cells in table 6.39 put in evidence how agent of this kind are massed in these communities (only in these five communities they have a double-digit presence). Thus, even if the contingency table regarding the typology of agents does not reveal clear elements of particular interest, as was observed also in the context of model BOOL_2, evidence has already emerged. In model ∂LoA the presence of groups of communities that show an intense presence of some specific typologies of agents is maybe clearer than in model BOOL_2.

### 6.3.4. Final considerations on DCI models

The characterizations that emerge also in model ∂LoA have to be observed with interest. From the beginning of BOOL_1 procedure, groups of agents distinguishable in terms of the typology of their agents have been highlighted. The first mask identified by the DCI algorithm in model BOOL_1 put in evidence a subset made up of 22 enterprises and 19 small local government institutions out of a total of 57 agents. It was clear that the methodology gave a first response in terms of agents' typologies, but there were no certainties about its ability to continue to produce results with the same orientation. However, this is what in fact happened, and the most interesting consideration, as for CPM and Infomap, regards the alignment between the purpose for which the methodology was applied and the results it produced. DCI was applied to investigate groups of agents with similar functions, and the detection of communities including agents characterized by specific typologies is the most meaningful result that could be expected. Since the methodology originated to investigate the presence of functional groups of agents, which in the case of Tononi *et al.* (1996, 1998) were neurons, the emergent conviction is that DCI algorithm produced these results because the nature of agents influence their functions and the function, that every agent has, determined its behavior. The use of DCI methodology in the context of this case study produced partitions in which similar agents tended to stay together probably because their common typology guided their activity (and thus, their participations) in funded projects. The conclusion is that the common nature of agents determined behavioral patterns that were the most integrated among all those considered. This, once again, has not to be confused with the presence of simultaneous and similar activities. The activities of agents belonging to one of the groups detected, were sometimes simultaneous and other times were not. Probably, most of the time it was an unintentional kind of coordination. However, an integration occurred and it is coherent to think that the typology affected and influenced it. Surely, the constraints of the waves were determinant in conditioning each time the agents' participations depending on their typologies, and maybe similar agents were more prone to communicate among themselves in order to decide whether to participate together or whether to alternate their presence. In any case, given the existing structure of relationships and given the presence of external constraints (which in this case were imposed by the public policies), the typology of agents seems to have determined specific functions which, in turn, inevitably influenced their behaviors. On the basis of considerations of this kind, the use of DCI analysis seems to have been perfectly coherent with the purpose for which it was selected: the detection of groups of agents that

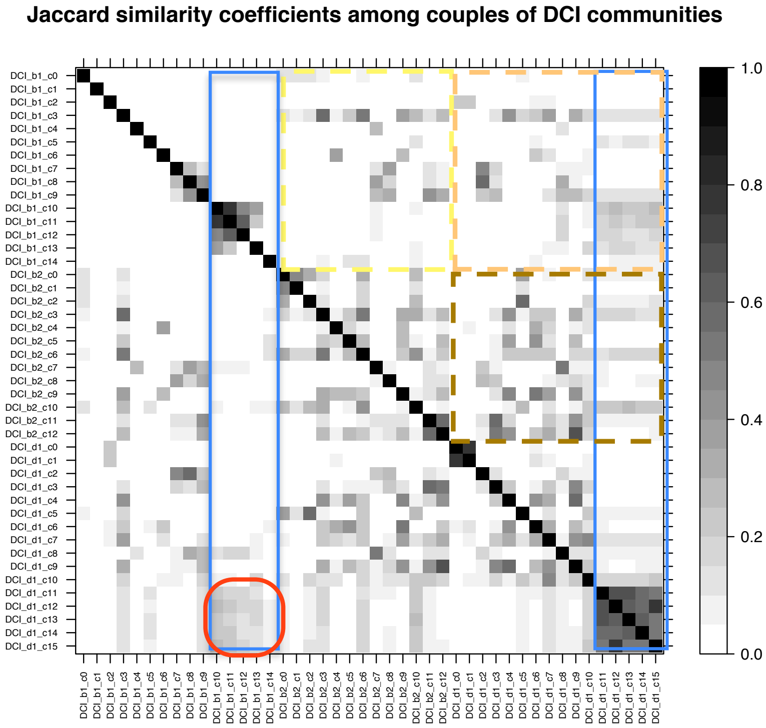



*Figure 6.41. Matrix of Jaccard similarity coefficients calculated for all pairs of communities detected in all three Infomapt models. The darker is the color in the cell, the higher is the ratio. The yellow dashed rectangle highlights combinations between BOOL_1 and BOOL_2 communities. Orange dashed rectangle highlights combinations between BOOL_1 and ∂LoA communities. Dark orange dashed rectangle highlights combinations between BOOL_2 and LoA communities. The two blue rectangles put in evidence the Jaccard similarity coefficients of two particular groups of communities. The red smoothed square highlights the coefficients determined by pairs of communities that correspond to the two blue rectangles.*

share similar functions. This is what the application of DCI highlighted in the specific case study and, besides producing interesting results, there were no previous analyses (concerning similar situations) that could be taken as point of reference. This fact contributes to enhance the importance of these results. As just chronicled, there were no previous examples or studies which could guarantee that such results could be achieved.

To conclude, a final consideration concerning the comparison among the three models developed. It is interesting to observe the graphical representation of the matrix of Jaccard similarity coefficients computed among all possible pairs of communities detected in each of the DCI models. Observing the figure 6.41, the two areas contoured by blue rectangles (the one on the left regards a group of communities detected in model BOOL_1 and the one on the right regards a group of communities detected in model ∂LoA) show communities which have intense and frequent overlaps among themselves but which, at the same time, do not have particular connections with other communities. However, it can be observed that the area in which they ‘meet’ (red smoothed square) seems to suggest that these two groups of communities have elements of similarity. This is interesting because the blue rectangle on the left identifies those communities that were detected through a final intense rearrangement of the system under analysis (round 11 in model BOOL_1). The model BOOL_2, even if using the same informational basis as model BOOL_1, did not give representation to these two groups which, as we know it from model ∂LoA, are strongly characterized by the presence of enterprises, KIBS and Universities/public research centers, and which thus have to be regarded with interest. This fact allows two final considerations. On one hand, the heuristic procedure that was adopted in the context of model BOOL_1 was able to group specific and important combinations of agents which the other model that used the same informational basis (model BOOL_2) was not able to detect. On the other hand, the use of a more detailed informational basis, as the one used in model ∂LoA (that regards the variation of levels of activity), provided an interesting grouping of agents using a stable and coherent procedure which, it has to be remembered, is affected by a strong disproportion (with respect to the previous analyses in which DCI was tested[98]) between the number of agents and the number of instants over which the system is observed. For sure, the possibility to observe the system under analysis over a larger number of instants will allow us to achieve more precise and consistent results but, nevertheless, what emerged from this work has to be regarded with interest.

## 6.4. Intersections among the communities detected through the different methodologies

The decision to apply different methodologies was taken after the identification of some key concepts in a specific theoretical framework concerning the study of innovation processes. The three elements that characterize the entities 'organizations' (*structure, processes and functions*) have been investigated coherently. This was the aim of this thesis, and the results described in the previous paragraphs of this chapter have to be considered the core of the development of the analysis. However, to conclude my work I wanted to check whether the different aspects investigated coexist and, if so, how this can be interpreted.
In order to investigate how the different methodologies produced intertwined results, I decided to study overlaps among communities detected through the different methodologies. To understand if there are agents that were grouped together by CPM, and by Infomap and by DCI analysis, I calculated the intersections generated by all possible combinations of three communities as follows:

- the first one is one of the communities detected with CPM;
- the second one is one of the communities detected with Infomap;
- the third one is one of the communities detected with DCI.

The intent was to find agents that were grouped together in at least one community detected in the context of every methodology applied, irrespectively of the specific model or the specific community in which they were found. Since CPM models produced 52 different communities, Infomap models 139 communities and DCI models 44 communities, the intersections of the 318.032 possible combinations were calculated. Out of them, 29.650 contain at least one agent and for each one of these I calculated the ratio between the

[98] See paragraphs 5.3.2 and 5.3.3

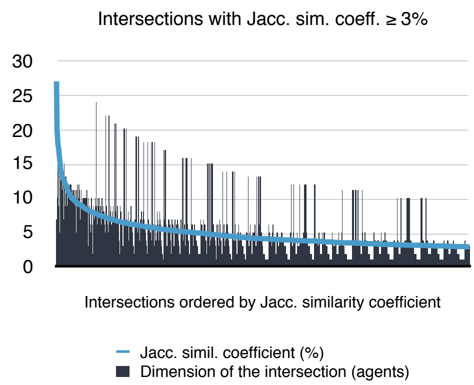



*Chart 6.42. On the X-axis are represented the intersections that have a Jaccard similarity ratio higher than or equal to 3.00%. The bars represent the dimensions (number of agents) of the intersections and the line indicates the Jacccard similarity ratio of the corresponding intersection detected.*

dimension (number of agents) of the intersection and the dimension (number of agents) of the union set. This measure, which is the Jaccard index, was used to exclude those in which the intersection constitutes a tiny part of the union set[99]. Fixing the threshold for the Jaccard index at a very low level (3,00%) made it possible to reduce the number of the intersections to 3.242 (that is 10,93 % of the non empty intersections). Even if information regarding each area of the overlap of the three communities considered was available (see figures 6.43a and 6.43b), my attention was focused exclusively on the simultaneous intersection. Thus, through R computation, I detected those combinations that generated perfectly equal intersections (in terms of the agents belonging to them). For every group of identical intersections I maintained only the one with the highest score in terms of the Jaccard index. In this way, I found 1.022 different intersections made up of agents[100] that were grouped together in (i) at least one of the communities produced by one of the CPM models, (ii) at least one of the communities produced by one of the Infomap models, and (iii) at least one of the communities produced by one of the DCI models[101].

Then, after having found agents that were grouped together in each one of the considered methodologies I attempted to have a final view concerning whether the simultaneous

[99] Moreover, the drop of those combinations in which the Jacquard coefficient was lower than 3.00% was also necessary since the algorithm I implemented in R to drop equal intersections (I describe it in this paragraph) cannot handle a big quantity of observations. Thus, a reduction was needed and it seemed to me appropriate to do it on the basis of the ratio between intersection and union set, rather than considering the dimension of one of them. This because it has not to be forgotten that the unit of analysis has to be retained the 'community'. Taking only a part of it does not guarantee that this portion would have been detected as a community. A subset of the agents that are in a specific 'organization' do not necessarily determine another 'organization'. Thus, to be the most coherent as possible with this last consideration, I decided to drop those combinations in which the intersection constitutes a tiny part of the union set.

[100] The agents involved in at least one intersection are 295 out of the 352 (those with at least two participations in funded projects) considered in the analysis.

[101] This procedure did not involve a preliminary exclusion of neither communities nor partitions detected. Even if every model allowed the detection of some partitions and/or communities that could be seen as more interesting than the others, the selection of some of them would have been an arbitrary procedure. Since at this stage of the work I had no specific reference and since the limitation of arbitrariness can be done only through further considerations the development of which is not a target of this work, I decided to proceed without any skimming of communities and/or partitions.

(a)

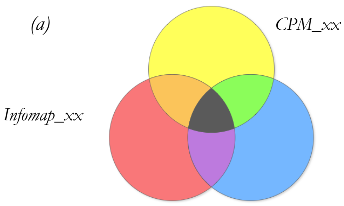



*Figure 6.43a (left). The Venn diagram represents the overlaps that were investigated. Every combination is made up of one community detected through CPM analysis, of one community detected through the application Infomap algorithm and of one community detected though DCI analysis. The overlap generates distinct 7 areas. CPM communities are colored in yellow, Infomap communities are colored in red, and DCI communities are colored in blue. The overlaps among them are colored in orange (CPM and Infomap), in purple (Infomap and DCI) and in green (DCI and CPM.), and the intersection of all three communities is colored in black.*

(b)

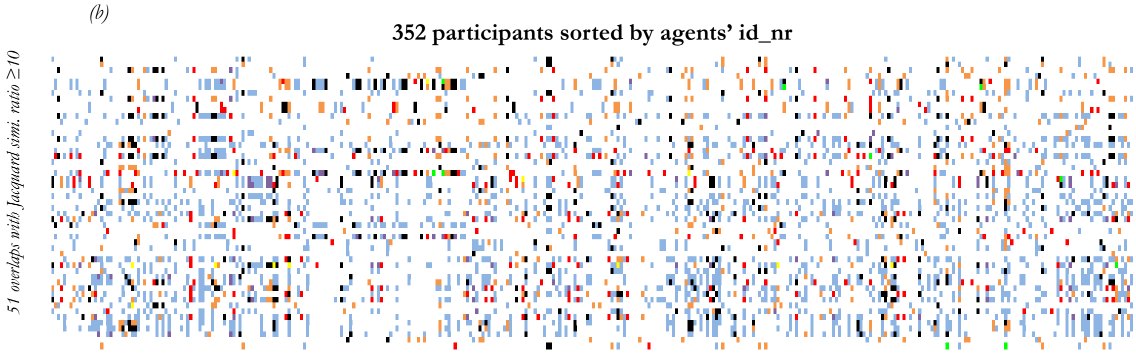



*Figure 6.43b (above). The matrix represents the 51 combinations with the a Jacccard similarity ratio higher than or equal to 10.00%. Each row represents a combination and each column represents an agent. Columns are ordered in descending order with respect of the corresponding Jaccard ratio (the one with the highest ration on top). Cells are colored in the same way as the Venn diagram, depending on the communities to which the agents belongs in that specific combination.*

(c)

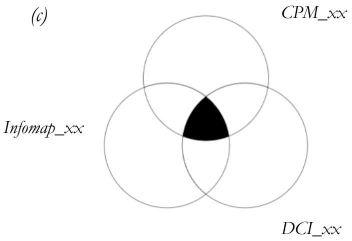
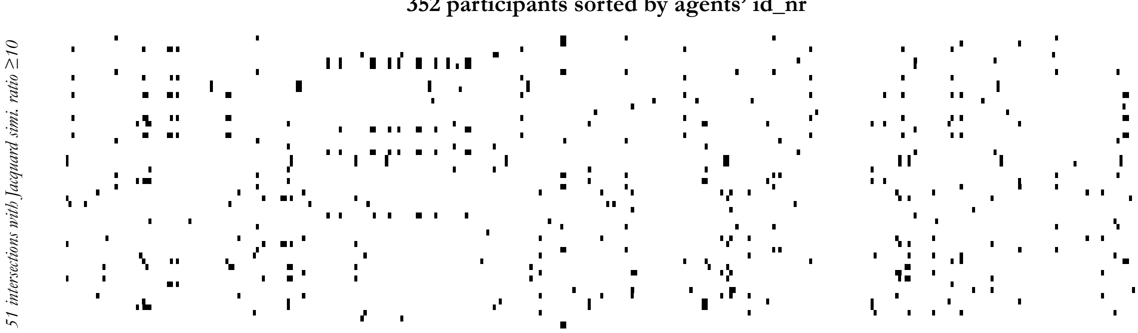



*Figure 6.43c (left). The figure represents the second step I made in the elaboration of the combinations of the overlaps determined by the combination of communities detected through the application of different methodologies. Only the intersections were taken into account.*

*Figure 6.43d (below). The matrix is the the same matrix represented in figure 6.42a but only intersections have been highlighted.*

(d)



presence of a *structure,* of *processes* and of *functions*[102] somehow characterized the activities of the agents involved. In order to do this, I merged the database with another database which I contributed to create in the context of another project in which I participated with the same research team. This project, which is called 'poli.in'[103], regarded the analysis and the modeling of the policy cycle that followed the one that was considered. After having supported innovation through the policies that have been taken into account in this work, in the programming period 2007-2013, the Italian region of Tuscany funded twelve 'innovation poles' with the aims (i) to strengthen the regional innovation system, (ii) to support the development of a range of knowledge-intensive services, and (iii) to encourage technology transfer and stimulate the innovation capabilities of regional small and medium-sized enterprises (SME). The main goal of this policy was slightly different from the preceding one. In fact, even if its general objective was again the sustaining of innovation, it involved the implementation of different kinds of measures that regarded the promotion of linkages between key regional actors (Universities, public research organizations, KIBS, large businesses and SME). Since the analysis that I developed was oriented to the investigation of how agents interacted among themselves in a specific context devoted to innovation, the best way I could attempt to make some final considerations was to observe how agents continued to have an active role in participating in innovation policies.
Thus, after the merge of the two databases (from it emerged that 199 agents out of the 352 considered in this work also participated in the following cycle of policies), I created a series of dummy variables starting from a series of continuos variables regarding the activities that the agents undertook in the policy cycle 2007-2013. The first dummy that I elaborated (d_net1) assumes value 1 if the corresponding agent, in the context of 2007-2013 policy, had at least one of these participations[104]:

(i) as managing organization of one of the innovation poles;
(ii) as agent affiliated to one of the innovation poles;
(iii) as services provider;
(iv) as services customer;
(v) as provider of laboratories (to the innovation poles);
(vi) as provider of workers (to the innovation poles).

The second dummy (d_net2) was elaborated in the same way but considering also if the agent had at least one of these participations:

(i) as managing organization of one of the innovation poles;

---

[102] Of course, the hypothesis present in this sentence can be considered quite strong since I am assuming that every community, depending on which was the methodology that allowed its detection, has a specific characterization in terms of what in the previous paragraph has been highlighted. This, without any doubt, is a generalization of the features emerged from the jointed observation of the partitions detected. However, since characterizations were funded and an interpretation was given, I decided to pursue the interpretation of the emerged aspects considering them in their broadest sense.

[103] Information about the project 'poli.in' is available at the web site www.poliinnovazione.unimore.it

[104] If only one of these conditions are true, the dummy 'd_net1' assumes value 1.

(ii) as partner in a collaboration agreement.

The idea was to counterpose the participation in two different kinds of networks. The first, the one to which variable 'd_net1' refers, regards substantially the activity in a context devoted to the supply/demand of resources (also the affiliation of agents to a pole can be related to the possibility to get some advantages in terms of services for their private activities). By contrast, the second network, the one to which variable 'd_net2' refers, is more related to the creation of common activities and the circulation of competences among agents.

I used these two dummies as independent variables in two regression models. Since my intent was not to develop in this context a model that can exactly predict the participation in the policy cycle 2007-2013, after having computed two different probit regressions with robust standard errors (the first with dependent variable 'd_net1' and the second with dependent variable 'd_net2'), I focused my attention exclusively on the statistical significance of the independent variables that I used, in order to understand if they contribute to explain the dependent variables and, if so, in which direction. These variables, that are meaningful in terms of what I have described in this paragraph, are:

i) the number of different intersections (determined as described above) in which the agent is present (N_int);
ii) the highest Jaccard ratio among those of the intersections in which the agent is present (Jacc_max);
iii) the dimension (in terms of number of agents) of the intersection (among those in which the agent is present) that has the highest Jaccard ratio (dim_int_maxJacc);
iv) a dummy for the typology of agents 'Universities and Public Research Centers' (d_R) and a dummy for the typology of agents 'Knowledge-Intensive Business Services' (d_KIBS)[105].

The results seem interesting. The first model shows a significant level only with respect of the dummy regarding the typology 'KIBS', and that is in line with the consideration of a network oriented to the supply/demand of resources. Moving to the second regression, something more emerged. The variables that reach significant levels are three: once again the dummy regarding the typology of agents KIBS (z=4,41), the dummy regarding Universities and Public Research Centers (z=5,34), and also the variable regarding the number of different intersections in which the agent is present (z=3,42). Also these results seem to be in line with the variable that is being investigated. Considering the participation in a network more devoted to supporting the circulation of competences, Universities and Public Research Centers have now a significant role and, along with them, also the number of different configurations (groups) that simultaneously were detected by the three different methodologies applied reaches a significant value.

Since some elements have emerged, I computed a third regression using again a probit model with robust standard errors. This time, the binary dependent variable (d_net3) was

[105] The introduction of these dummies was necessary since in the context of analysis of regional development policies these typologies of agents always have a key role.

| Probit model with Robust Standard Errors<br>d_net1 as dependent variable<br>(in blue z≥1,96) | | | |
|---|---|---|---|
| **Independent Variables** | **Coeff.** | **z** | **P-value** |
| **N_int_abvTHR** | 0,004793 | 1,49 | 0,137 |
| **jacc_max** | -0,019687 | -1,56 | 0,118 |
| **dim_int_with_maxJacc** | 0,014825 | 1,24 | 0,215 |
| **d_R** | 0,053111 | 0,20 | 0,840 |
| **d_KIBS** | 0,302392 | 2,02 | 0,044 |

| Probit model d_net2 with Robust Standard Errors<br>d_net2 as dependent variable<br>(in blue z≥1,96) | | | |
|---|---|---|---|
| **Independent Variables** | **Coeff.** | **z** | **P-value** |
| **N_int_abvTHR** | 0,013303 | 3,42 | 0,001 |
| **jacc_max** | 0,021782 | 1,25 | 0,210 |
| **dim_int_with_maxJacc** | -0,001673 | -0,10 | 0,920 |
| **d_R** | 1,689695 | 5,34 | 0,000 |
| **d_KIBS** | 1,064061 | 4,41 | 0,000 |

*Table 6.44 (left). Results of the probit regression with dummy d_net1 as dependent variable. Cells with a z statistics higher than or equal to 1,96 are colored in blue.*

*Table 6.45 (right). Results of the probit regression with dummy d_net2 as dependent variable. Cells with a z statistics higher than or equal to 1,96 are colored in blue.*

defined equal to 1 only in the case that the corresponding agent participated in at least one collaboration agreement, which above all has to be considered the type of relation most oriented to the sharing of information and to the common purpose to develop innovative activities. The set of the independent variables was always the same as the preceding regressions, and the model showed an emerging interesting result. The variables that have a statistic higher than or equal to 1,96 are the same as the preceding model (d_net2) and, in addition, one more variable now appears as significant: the variable describing the maximum Jaccard ratio. The presence of the agent in an 'intense' intersection seems to be related to its participation in networks devoted to the activation of new interactive processes. This is interesting since, if the three communities (each of which was intended to represent a different aspect of the entity 'organization') generate an intersection characterized by a dimension that is commensurate with the corresponding union set, what seems to appear is a situation of systemic involvement of the agents. In other words, the higher the Jaccard ratio, the better the agents involved in the intersection represent the totality of the three communities, and the observation of a statistical significance of this variable seems to highlight that the participation in an 'intense organization' has a role in the development of new interacting activities fostering the sharing of competences.

*Table 6.46 (right). Results of the probit regression with dummy d_net3 as dependent variable. Cells with a z statistics higher than or equal to 1.96 are colored in blue.*

| Probit model d_net3 with Robust Standard Errors<br>d_net3 as dependent variable<br>(in blue z≥1,96) | | | |
|---|---|---|---|
| **Independent Variables** | **Coeff.** | **z** | **P-value** |
| **N_int_abvTHR** | 0,015421 | 3,70 | 0,000 |
| **jacc_max** | 0,041187 | 2,18 | 0,029 |
| **dim_int_with_maxJacc** | -0,020345 | -1,09 | 0,277 |
| **d_R** | 1,914026 | 5,25 | 0,000 |
| **d_KIBS** | 0,953101 | 3,25 | 0,001 |

Without doubt, these results have to be better investigated with further and more in-depth research[106]. Nevertheless, to conclude this thesis, I felt it was important to bring back to a single path the processes of detection of the three different aspects of the idea of 'organization' within a theoretical framework analyzing innovation, and to attempt to provide some essential flashes for further developments. After having considered different methodologies that allowed the investigation of crucial aspects, the description of how the different communities could be combined together has to be intended only as a creative proposal aimed at making possible some final considerations. I focused my attention on the agents that are viewed together as a group from each of the three different perspectives considered, and the observation of how they continued their activities in the following public policy cycle, allows us to glimpse some interesting results which seem to be in line with the hypotheses from which the whole work initiated.

[106] Of course results have to be better investigated with further and more in-depth analyses in terms of omitted variables, coefficients but also in accuracy, but at the same time what emerged has to be considered with interest.

# Conclusion

The analysis of the cycle of polices that Region Tuscany promoted from 2000 to 2006, in order to support the formation of innovation processes among local actors, was developed through the application of three different methodologies. Each of them was selected after the identification of three crucial aspects that David Lane's theory of innovation (Lane and Maxfield 1997, 2005; Lane 2011) highlights as determinant elements with regard to interacting agents aiming to foster innovation processes. The presence of connective structures, of shared processes and of common functions, was investigated with the intent to understand if the observed interactions could have determined the formation of innovative communities that go beyond the boundaries of the partnerships constituted for the realization of the single projects. First of all, I applied Clique Percolation Method (CPM) by Palla *et al.* (2005) in order to investigate how the common participations in funded projects determined patterns of connections that can suggest the presence of communities. Then, I applied Infomap algorithm by Rosvall and Bergstrom (2008) to study, through different constrained simulations of flows over the network, the processes that occurred in the context of the considered policies. Finally, I applied the Dynamic Cluster Index analysis (DCI) by Villani, Filisetti, Benedettini, Roli, Lane and Serra (2013), a new methodology which, focusing on the integration of agents' behaviors, allows the identification of functional subsystems without using any kind of information regarding the topology of the network.

For each one of these methodologies, different models were developed. First of all, in the context of CPM analysis, the problematic definition of the value of $k$ was overcome thanks to the observation of the specific features of all possible partitions that the algorithm makes it possible to detect. Thus, after the identification of three ranges of $k$, three specific values of it were selected ($k$=5, $k$=12 and $k$=18) and three models were elaborated. Then, in the context of Infomap analysis I computed simulations through the development of (i) a model characterized by a memoryless flow, and (ii) two second order Markov models elaborated considering the chronological sequence in which projects were carried out. Finally, in the context of DCI methodology I developed its application to a socio-economic context of analysis. An explorative process was determinant to make emerge some limits, mostly related to the available informational basis, which required the evaluation of new strategies. Then, after having elaborated specific proposals, I developed two other processes of analysis over two different sets of variables describing the behavioral profiles of agents.

The tailoring of these nine models required in-depth understanding of the methodologies considered and also an intense degree of originality, since no specific literature references were found concerning previous applications to similar case studies. All these nine models were developed taking into account the peculiarities of the object under analysis: (i) a cycle of innovation network policies made up of nine waves with various important features, (ii) in which agents' participations were not constant and (iii) in which partnerships' compositions changed over time. The complex nature of the processes supported (innovative projects) and the discontinuity of temporal dynamics have to be borne in mind in order to understand how the different methodologies and the different models were considered and developed in an intricate context of analysis.

The results that emerged from the different analyses seem to be in line with what was investigated. Even if further research is needed in order to better understand how agents' participations in innovative projects can have affected the creation of a solid and dynamic interactive pattern, the analyses developed revealed meaningful partitions. First of all, the application of the three different models of CPM allowed the detection of communities that the higher the values of *k*, the higher their characterization in terms of participations in specific waves. The observation of meaningful contingency tables allowed me to see how, especially in the model with *k* equal to 18, every detected group is characterized by a significant number of participations in projects developed in one of the nine waves. Thus, having applied CPM to study the presence of communities characterized by intense relational structures, it seems to me coherent that the detected groups are made up of agents that have significant participations of this kind. The waves and, before them, the policies' implementation, were crucial to determine the shape of the agents' connective framework. Through the definition of specific domains of interventions and through the imposition of different constraints to the agents' participations, the policies that were implemented over time necessarily affected the construction of the network relations.

Infomap models also show results in line with the purpose for which this methodology was applied. The three partitions detected are made up of communities characterized by agents' participations in projects related to specific technological domains. A similarity emerged between the communities identified, especially those detected by imposing restrictions on the circulation of the flows, and the partnerships that during the policies were realized. Infomap methodology, which was applied to investigate groups of agents characterized by the capacity to share working processes, allowed me to detect communities which seem to reflect the participations in the most important common activities observed in the context of the cycle of policies: the projects. A further investigation of how these groups proceeded (or did not) to develop shared activities would probably provide useful elements in the comprehension of how the policies really supported the creation of continuative innovative collaborations.

Finally, the application of DCI algorithm, the methodology to which I dedicated the most of my efforts during the last year of my Ph.D., also revealed results that deserve attention. The partitions that were detected were analyzed through the computation of meaningful contingency tables. While with the preceding methodologies characterizations regarding the agents' participations emerged, with this analysis the communities detected seem to be related to a different kind of feature, the typology of the agents involved. The third model in particular, the one in which the method was applied over the most refined informational basis and in which I introduced a procedure of cluster analysis of the large quantity of masks detected by the algorithm, allowed the identification of a partition in which, on one hand, KIBS and enterprises, and on the other, KIBS and enterprises and universities, determine two distinct groups of communities. Since this methodology originated to detect functional groups (Tononi *et al.* 1994), the emergence of the salience of the typology of agents seems to underline that similar agents have similar finalities. The presence of integrated activities (that are what DCI algorithm investigates) can be reasonably related to the presence of similar typologies of economic institutions, since it can be said that the

specific nature of agents necessarily influences their functions and these, in turn, determine specific behavioral patterns.
This is what the application of the different methodologies produced and what their results suggested. This work has attempted to start out on possible paths to analyze a complex theme, that of innovative organizations, and clearly the results have to be further investigated. Nevertheless, some elements have emerged and they seem to suggest that these analyses were coherent with the intent to which they were applied. Since there are no specific examples in the literature that testify to the appropriacy of the application of these methodologies to the investigation of innovative dynamics in a socio-economic context, there was no guarantee that the described characterizations would be found.
To conclude the work, in order to bring back to a single route the three paths on which I embarked during this thesis, I analyzed the intersections generated by the overlaps of the communities detected through the different methodologies. The detection of agents that are viewed together from each of the three perspectives assumed (that of structure, that of processes and that of functions) have revealed some emergent characterizations in terms of the participations in the following policy cycle that Region Tuscany developed in support of innovation. This last part has to be considered only as a proposal, while the core of the work finds its place in the attempt to maintain coherence between the way the analyses were developed and the purposes for which they were applied.

# Appendix

## a. Contingency tables and graphs of the communities detected

### a.1. CPM_k05

*Typology of agents (obs. freq)*

| CPM_k05 | GI | TA | CC | SC | LGI | E | R | BS | KIBS |
|---|---|---|---|---|---|---|---|---|---|
| C0 | 1 | | | | 1 | 2 | | | 1 |
| C1 | | | | | | 1 | | | 5 |
| C2 | 11 | 7 | 10 | 2 | 38 | 91 | 27 | 5 | 107 |
| C3 | | | | | | 1 | 1 | | 3 |
| C4 | 1 | | | | 2 | 3 | | | 2 |
| C5 | 1 | | | | | | 1 | 1 | 2 |
| C6 | | | 1 | | | 2 | 1 | | 4 |
| C7 | | | | | 1 | 4 | 1 | | 2 |
| C8 | | | | | | | 4 | | 1 |
| C9 | | | | 1 | | 1 | | | 3 |
| C10 | | | | | | 3 | | | 3 |
| C11 | | | | | 2 | 3 | 1 | | 4 |
| C12 | 1 | 2 | | | | 4 | 2 | | 3 |
| C13 | 3 | | | | 5 | 6 | 1 | 1 | 2 |
| C14 | | | | | | 1 | 1 | | 3 |

*Typology of agents (adj. res)*

| CPM_k05 | GI | TA | CC | SC | LGI | E | R | BS | KIBS |
|---|---|---|---|---|---|---|---|---|---|
| C0 | 1,70 | -0,34 | -0,38 | -0,20 | 0,54 | 0,48 | -0,75 | -0,30 | -0,75 |
| C1 | -0,53 | -0,37 | -0,41 | -0,21 | -0,92 | -0,73 | -0,82 | -0,33 | 2,44 |
| C2 | -1,25 | 0,28 | 1,31 | -0,28 | 0,64 | 0,25 | -0,95 | -0,14 | 0,01 |
| C3 | -0,49 | -0,34 | -0,38 | -0,20 | -0,84 | -0,50 | 0,76 | -0,30 | 1,13 |
| C4 | 1,11 | -0,43 | -0,48 | -0,25 | 1,13 | 0,45 | -0,95 | -0,38 | -0,65 |
| C5 | 1,70 | -0,34 | -0,38 | -0,20 | -0,84 | -1,48 | 0,76 | 3,15 | 0,19 |
| C6 | -0,62 | -0,43 | 1,72 | -0,25 | -1,06 | -0,32 | 0,25 | -0,38 | 0,84 |
| C7 | -0,62 | -0,43 | -0,48 | -0,25 | 0,03 | 1,23 | 0,25 | -0,38 | -0,65 |
| C8 | -0,49 | -0,34 | -0,38 | -0,20 | -0,84 | -1,48 | 5,28 | -0,30 | -0,75 |
| C9 | -0,49 | -0,34 | -0,38 | 5,05 | -0,84 | -0,50 | -0,75 | -0,30 | 1,13 |
| C10 | -0,53 | -0,37 | -0,41 | -0,21 | -0,92 | 1,06 | -0,82 | -0,33 | 0,73 |
| C11 | -0,69 | -0,48 | -0,54 | -0,28 | 0,77 | -0,01 | 0,01 | -0,43 | 0,27 |
| C12 | 0,66 | 3,44 | -0,59 | -0,30 | -1,31 | 0,24 | 0,80 | -0,47 | -0,80 |
| C13 | 2,57 | -0,66 | -0,73 | -0,38 | 2,08 | 0,30 | -0,63 | 1,27 | -2,24 |
| C14 | -0,49 | -0,34 | -0,38 | -0,20 | -0,84 | -0,50 | 0,76 | -0,30 | 1,13 |

*Participations of agents in waves (obs. freq)*

| CPM_k05 | 2002_ 171 | 2002_ 172 | 2002_ ITT | 2004_ 171 | 2004_ 171A | 2005_ 171 | 2006_ VIN | 2007_ 171 | 2008_ 171 |
|---|---|---|---|---|---|---|---|---|---|
| C0 | | | 8 | | | 5 | 2 | 4 | |
| C1 | | 3 | | | | 11 | 7 | | |
| C2 | 117 | 69 | 300 | 62 | 55 | 457 | 10 | 43 | 2 |
| C3 | 4 | 17 | | | | | | | |
| C4 | | 22 | 6 | | | 8 | | | |
| C5 | | | 17 | 16 | | | | | |
| C6 | 25 | | 19 | 12 | | | | | |
| C7 | 19 | 14 | | | | 5 | | 6 | |
| C8 | | | 39 | | | | 2 | 2 | |
| C9 | 23 | | | | 8 | 2 | | 3 | |
| C10 | | | 21 | 4 | 2 | | 3 | 3 | |
| C11 | 24 | 8 | 12 | | 7 | 4 | | | |
| C12 | 19 | | 38 | | | | | | |
| C13 | 3 | 6 | 12 | 4 | | 38 | | | |
| C14 | | 8 | | | | 11 | | | |

*Participations of agents in waves (adj. res.)*

| CPM_k05 | 2002_ 171 | 2002_ 172 | 2002_ ITT | 2004_ 171 | 2004_ 171A | 2005_ 171 | 2006_ VIN | 2007_ 171 | 2008_ 171 |
|---|---|---|---|---|---|---|---|---|---|
| C0 | -1,78 | -1,37 | 1,31 | -1,10 | -0,94 | -0,60 | 3,32 | 4,03 | -0,15 |
| C1 | -1,87 | 0,87 | -2,92 | -1,16 | -0,99 | 1,93 | 12,28 | -0,90 | -0,16 |
| C2 | -6,18 | -5,59 | -2,18 | -0,93 | 1,64 | 10,26 | -2,73 | 0,50 | 0,98 |
| C3 | 0,65 | 11,67 | -2,92 | -1,16 | -0,99 | -3,22 | -0,56 | -0,90 | -0,16 |
| C4 | -2,47 | 11,12 | -1,60 | -1,52 | -1,30 | -1,36 | -0,74 | -1,19 | -0,21 |
| C5 | -2,36 | -1,81 | 2,94 | 10,45 | -1,24 | -4,05 | -0,71 | -1,14 | -0,20 |
| C6 | 6,65 | -2,38 | 0,90 | 4,99 | -1,63 | -5,32 | -0,93 | -1,49 | -0,27 |
| C7 | 5,59 | 5,41 | -4,25 | -1,69 | -1,44 | -3,07 | -0,82 | 3,54 | -0,23 |
| C8 | -2,70 | -2,08 | 9,13 | -1,67 | -1,42 | -4,64 | 1,78 | 0,34 | -0,23 |
| C9 | 8,65 | -1,90 | -3,84 | -1,52 | 5,31 | -3,52 | -0,74 | 1,49 | -0,21 |
| C10 | -2,36 | -1,81 | 4,50 | 1,52 | 0,48 | -4,05 | 3,70 | 1,66 | -0,20 |
| C11 | 6,37 | 1,49 | -1,13 | -1,90 | 3,09 | -4,10 | -0,92 | -1,48 | -0,26 |
| C12 | 4,22 | -2,40 | 6,48 | -1,93 | -1,64 | -5,36 | -0,93 | -1,51 | -0,27 |
| C13 | -2,18 | 0,18 | -1,71 | 0,14 | -1,73 | 4,75 | -0,98 | -1,59 | -0,28 |
| C14 | -1,78 | 5,11 | -2,77 | -1,10 | -0,94 | 2,35 | -0,53 | -0,86 | -0,15 |

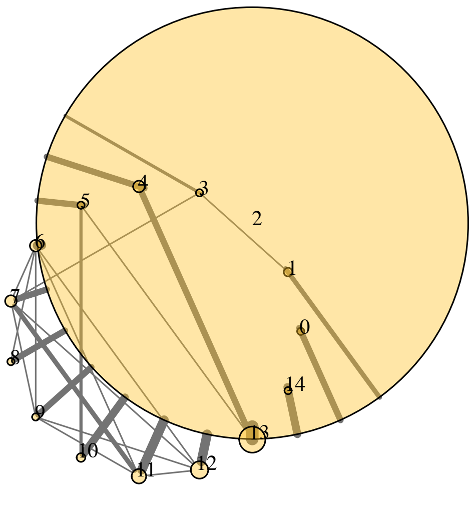
*Participations of agents in projects with specific technological domains (obs. freq)*

| CPM_k05 | Others | Bio-Technologies | Organic Chemistry | Geotermal Sciences and Biomasses | ICT - Multimedia | Mechanics Engineering | Multi Disciplinary | Nanotechnologies | New Materials | Optoelectronic | NA |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **C0** | | | | | 11 | | | | | 8 | 0 |
| **C1** | | 3 | | | 18 | | | | | 0 | 0 |
| **C2** | 4 | 97 | 48 | 30 | 592 | 92 | 13 | 39 | 33 | 153 | 14 |
| **C3** | | 17 | | | 4 | | | | | 0 | 0 |
| **C4** | | | | | 30 | | | | | 6 | 0 |
| **C5** | | 17 | | 16 | | | | | | 0 | 0 |
| **C6** | | | 12 | | 41 | | | 3 | | 0 | 0 |
| **C7** | 12 | 10 | | | 22 | | | | | 0 | 0 |
| **C8** | | | | | 23 | | | | | 18 | 2 |
| **C9** | | | 11 | | 25 | | | | | 0 | 0 |
| **C10** | | 17 | 3 | 6 | 3 | | | 4 | | 0 | 0 |
| **C11** | 12 | | | 7 | 36 | | | | | 0 | 0 |
| **C12** | | 34 | 2 | | 21 | | | | | 0 | 0 |
| **C13** | | 6 | | 4 | 47 | | | | | 6 | 0 |
| **C14** | | | | | 19 | | | | | 0 | 0 |

*Participations of agents in projects with specific technological domains (adj. res.)*

| CPM_k05 | Others | Bio-Technologies | Organic Chemistry | Geotermal Sciences and Biomasses | ICT - Multimedia | Mechanics Engineering | Multi Disciplinary | Nanotechnologies | New Materials | Optoelectronic | NA |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **C0** | -0,58 | -1,63 | -0,96 | -0,87 | 0,34 | -1,07 | -0,39 | -0,74 | -0,63 | 4,19 | -0,43 |
| **C1** | -0,61 | 0,30 | -1,01 | -0,92 | 2,93 | -1,12 | -0,41 | -0,78 | -0,66 | -1,67 | -0,46 |
| **C2** | -6,07 | -6,23 | -0,83 | -3,44 | -1,10 | 6,84 | 2,51 | 2,53 | 4,02 | 3,95 | 1,71 |
| **C3** | -0,61 | 9,70 | -1,01 | -0,92 | -3,24 | -1,12 | -0,41 | -0,78 | -0,66 | -1,67 | -0,46 |
| **C4** | -0,80 | -2,26 | -1,33 | -1,21 | 3,57 | -1,47 | -0,54 | -1,03 | -0,87 | 0,97 | -0,60 |
| **C5** | -0,76 | 6,98 | -1,28 | 13,53 | -6,29 | -1,41 | -0,52 | -0,98 | -0,83 | -2,10 | -0,57 |
| **C6** | -1,00 | -2,84 | 6,11 | -1,52 | 2,93 | -1,85 | -0,68 | 1,19 | -1,09 | -2,75 | -0,75 |
| **C7** | 13,32 | 2,17 | -1,48 | -1,34 | -0,54 | -1,63 | -0,60 | -1,14 | -0,96 | -2,43 | -0,67 |
| **C8** | -0,87 | -2,47 | -1,46 | -1,32 | -0,07 | -1,61 | -0,59 | -1,13 | -0,95 | 6,29 | 2,50 |
| **C9** | -0,80 | -2,26 | 7,51 | -1,21 | 1,88 | -1,47 | -0,54 | -1,03 | -0,87 | -2,19 | -0,60 |
| **C10** | -0,76 | 6,98 | 1,24 | 4,35 | -5,23 | -1,41 | -0,52 | 3,29 | -0,83 | -2,10 | -0,57 |
| **C11** | 11,76 | -2,81 | -1,66 | 3,51 | 1,73 | -1,83 | -0,67 | -1,28 | -1,08 | -2,73 | -0,75 |
| **C12** | -1,01 | 11,16 | -0,40 | -1,53 | -2,65 | -1,87 | -0,69 | -1,30 | -1,10 | -2,78 | -0,76 |
| **C13** | -1,06 | -0,66 | -1,78 | 1,07 | 3,34 | -1,97 | -0,72 | -1,37 | -1,16 | -0,52 | -0,80 |
| **C14** | -0,58 | -1,63 | -0,96 | -0,87 | 4,04 | -1,07 | -0,39 | -0,74 | -0,63 | -1,59 | -0,43 |

**CPM_k05**




*Graph of the partition. Node are proportional to the number of agents that belong to the corresponding community. Edges represent overlaps between communities and they are proportional to the number of nodes involved. Circle layout.*

## a.2. CPM_k12

*Typology of agents (obs. freq)*

| CPM_k12 | GI | TA | CC | SC | LGI | E | R | BS | KIBS |
|---|---|---|---|---|---|---|---|---|---|
| C0 | 1 | | | | | 6 | 1 | | 4 |
| C1 | | | 1 | | 2 | 4 | 3 | | 4 |
| C2 | | | | | 2 | 3 | 1 | | 6 |
| C3 | 1 | | | | 1 | 4 | 1 | | 6 |
| C4 | | | 1 | | 2 | 2 | | | 9 |
| C5 | | | 1 | | 3 | 1 | 2 | | 6 |
| C6 | 1 | 1 | | | 4 | 2 | 1 | | 5 |
| C7 | | | | | 2 | 2 | 1 | | 7 |
| C8 | 3 | | | | 5 | 6 | 1 | 1 | 1 |
| C9 | | 1 | | | 4 | 4 | 1 | | 6 |
| C10 | 1 | | 1 | | 2 | 4 | 1 | | 7 |
| C11 | 3 | 1 | 2 | | 4 | 2 | 1 | 1 | 11 |
| C12 | 2 | | | | 2 | | 1 | 1 | 6 |
| C13 | 2 | | | | 1 | 1 | 2 | 1 | 5 |
| C14 | | 1 | 2 | 1 | 2 | 9 | 4 | 1 | 14 |
| C15 | 1 | | 1 | | 5 | | 2 | | 11 |
| C16 | | | | | 1 | 6 | 1 | | 6 |
| C17 | 1 | | | | 4 | 8 | 1 | | 8 |
| C18 | | | 2 | | 2 | 12 | | | 5 |
| C19 | | | | | 3 | 1 | 2 | | 11 |
| C20 | 1 | | | | | 3 | 1 | | 8 |
| C21 | 1 | 2 | 1 | | 1 | 4 | 1 | | 5 |
| C22 | 1 | 2 | | | 11 | 9 | 1 | | 14 |
| C23 | | | 2 | | 5 | 2 | 4 | 1 | 6 |
| C24 | | | 1 | | 2 | 4 | 3 | | 14 |
| C25 | | | | 1 | | 2 | 1 | 1 | 15 |
| C26 | 1 | | 2 | | 4 | | 1 | | 8 |

*Typology of agents (adj. res)*

| CPM_k12 | GI | TA | CC | SC | LGI | E | R | BS | KIBS |
|---|---|---|---|---|---|---|---|---|---|
| C0 | 0,72 | -0,46 | -0,68 | -0,23 | -1,51 | 2,47 | 0,02 | -0,43 | -0,73 |
| C1 | -0,80 | -0,50 | 0,73 | -0,25 | -0,13 | 0,68 | 1,83 | -0,46 | -1,16 |
| C2 | -0,74 | -0,46 | -0,68 | -0,23 | 0,11 | 0,33 | 0,02 | -0,43 | 0,45 |
| C3 | 0,64 | -0,48 | -0,70 | -0,24 | -0,79 | 0,85 | -0,07 | -0,45 | 0,18 |
| C4 | -0,80 | -0,50 | 0,73 | -0,25 | -0,13 | -0,64 | -1,14 | -0,46 | 1,58 |
| C5 | -0,77 | -0,48 | 0,81 | -0,24 | 0,76 | -1,21 | 0,96 | -0,45 | 0,18 |
| C6 | 0,56 | 1,61 | -0,73 | -0,25 | 1,37 | -0,64 | -0,15 | -0,46 | -0,61 |
| C7 | -0,74 | -0,46 | -0,68 | -0,23 | 0,11 | -0,39 | 0,02 | -0,43 | 1,04 |
| C8 | 2,81 | -0,55 | -0,81 | -0,27 | 1,61 | 1,45 | -0,35 | 1,54 | -3,20 |
| C9 | -0,85 | 1,45 | -0,78 | -0,26 | 1,06 | 0,38 | -0,29 | -0,50 | -0,51 |
| C10 | 0,42 | -0,53 | 0,59 | -0,26 | -0,34 | 0,38 | -0,29 | -0,50 | |
| C11 | 2,00 | 0,93 | 1,23 | -0,33 | 0,06 | -1,66 | -0,79 | 1,08 | 0,03 |
| C12 | 2,18 | -0,46 | -0,68 | -0,23 | 0,11 | -1,82 | 0,02 | 2,00 | 0,45 |
| C13 | 2,18 | -0,46 | -0,68 | -0,23 | -0,70 | -1,11 | 1,08 | 2,00 | -0,14 |
| C14 | -1,27 | 0,59 | 0,75 | 2,36 | -1,61 | 0,78 | 0,79 | 0,74 | -0,31 |
| C15 | 0,18 | -0,60 | 0,35 | -0,30 | 1,19 | -2,37 | 0,30 | -0,56 | 1,04 |
| C16 | -0,80 | -0,50 | -0,73 | -0,25 | -0,88 | 2,01 | -0,15 | -0,46 | -0,06 |
| C17 | 0,08 | -0,63 | -0,92 | -0,31 | 0,35 | 1,78 | -0,64 | -0,59 | -0,71 |
| C18 | -0,98 | -0,61 | 1,50 | -0,30 | -0,78 | 4,12 | -1,40 | -0,57 | -1,88 |
| C19 | -0,88 | -0,55 | -0,81 | -0,27 | 0,24 | -1,58 | 0,55 | -0,51 | 1,78 |
| C20 | 0,64 | -0,48 | -0,70 | -0,24 | -1,57 | 0,17 | -0,07 | -0,45 | 1,32 |
| C21 | 0,48 | 3,57 | 0,66 | -0,26 | -0,96 | 0,52 | -0,22 | -0,48 | -0,82 |
| C22 | -0,50 | 1,79 | -1,24 | -0,42 | 2,38 | 0,39 | -1,30 | -0,79 | -0,89 |
| C23 | -0,96 | -0,60 | 1,58 | -0,30 | 1,19 | -1,25 | 1,97 | 1,34 | -1,26 |
| C24 | -1,05 | -0,66 | 0,16 | -0,33 | -1,00 | -0,56 | 0,79 | -0,61 | 1,48 |
| C25 | -0,96 | -0,60 | -0,88 | 3,24 | -1,96 | -1,25 | -0,53 | 1,34 | 2,88 |
| C26 | 0,42 | -0,53 | 1,96 | -0,26 | 1,06 | -2,11 | -0,29 | -0,50 | 0,52 |

*Participations of agents in waves (obs. freq)*

| CPM_k12 | 2002_171 | 2002_172 | 2002_ITT | 2004_171 | 2004_171A | 2005_171 | 2006_VIN | 2007_171 |
|---|---|---|---|---|---|---|---|---|
| C0 | | | 71 | | | 3 | 4 | |
| C1 | 11 | 10 | 45 | | 14 | 24 | | |
| C2 | 3 | 8 | | | | 37 | | |
| C3 | | | 77 | | | | | |
| C4 | 8 | | 24 | | 8 | 37 | | |
| C5 | | 5 | 35 | | 9 | 28 | | |
| C6 | 8 | 8 | | | | 38 | | |
| C7 | | 5 | 34 | | | 46 | | |
| C8 | 3 | 6 | 6 | 4 | | 38 | | |
| C9 | 22 | 10 | 14 | 2 | 17 | 28 | | |
| C10 | 30 | | 13 | | | 62 | | |
| C11 | 12 | 5 | 32 | | 15 | 72 | | |
| C12 | | | 12 | | 6 | 49 | | |
| C13 | | | 15 | | 6 | 22 | | 2 |
| C14 | 39 | 12 | 42 | 27 | | 33 | | 8 |
| C15 | | 7 | 8 | 8 | 15 | 64 | | |
| C16 | | | 79 | | | 16 | | |
| C17 | 19 | 4 | 30 | | | 79 | | |
| C18 | 31 | | 2 | 33 | | 19 | | |
| C19 | | 9 | 30 | 5 | | 67 | | |
| C20 | | | 87 | | | | | |
| C21 | 5 | | 25 | | 4 | 31 | | |
| C22 | 38 | 8 | 26 | | 7 | 109 | | |
| C23 | 18 | 10 | 50 | | 15 | 56 | | |
| C24 | | 13 | 65 | 5 | | 81 | | |
| C25 | 117 | | | | | | | |
| C26 | 5 | | 13 | 4 | 9 | 43 | | |

*Participations of agents in waves (adj. res.)*

| CPM_k12 | 2002_171 | 2002_172 | 2002_ITT | 2004_171 | 2004_171A | 2005_171 | 2006_VIN | 2007_171 |
|---|---|---|---|---|---|---|---|---|
| C0 | -3,62 | -1,96 | 11,43 | -1,67 | -2,00 | -6,79 | 11,46 | -0,55 |
| C1 | -1,03 | 2,52 | 2,58 | -1,94 | 4,26 | -3,81 | -0,41 | -0,64 |
| C2 | -1,56 | 4,06 | -4,77 | -1,30 | -1,56 | 5,12 | -0,27 | -0,43 |
| C3 | -3,60 | -1,95 | 13,07 | -1,66 | -1,99 | -7,44 | -0,35 | -0,55 |
| C4 | -0,93 | -1,95 | -0,10 | -1,66 | 2,36 | 1,26 | -0,35 | -0,55 |
| C5 | -3,60 | 0,83 | 2,63 | -1,66 | 2,91 | -0,86 | -0,35 | -0,55 |
| C6 | 0,17 | 3,65 | -5,06 | -1,38 | -1,66 | 4,42 | -0,29 | -0,46 |
| C7 | -3,78 | 0,60 | 1,67 | -1,74 | -2,09 | 2,48 | -0,37 | -0,58 |
| C8 | -1,92 | 2,18 | -3,48 | 1,56 | -1,70 | 3,97 | -0,30 | -0,47 |
| C9 | 2,73 | 2,92 | -3,52 | -0,65 | 6,25 | -2,19 | -0,38 | -0,61 |
| C10 | 4,39 | -2,29 | -4,34 | -1,95 | -2,34 | 3,82 | -0,41 | -0,65 |
| C11 | -1,79 | -0,51 | -2,11 | -2,23 | 3,54 | 2,88 | -0,47 | -0,74 |
| C12 | -3,35 | -1,81 | -2,46 | -1,54 | 1,64 | 5,40 | -0,32 | -0,51 |
| C13 | -2,73 | -1,48 | 0,24 | -1,26 | 2,73 | 1,07 | -0,26 | 4,47 |
| C14 | 3,85 | 1,82 | -1,58 | 9,78 | -2,92 | -5,48 | -0,51 | 9,77 |
| C15 | -4,16 | 1,14 | -5,28 | 2,58 | 4,82 | 4,53 | -0,40 | -0,64 |
| C16 | -4,01 | -2,17 | 10,98 | -1,85 | -2,22 | -4,89 | -0,39 | -0,61 |
| C17 | 0,13 | -0,86 | -2,28 | -2,19 | -2,63 | 4,49 | -0,46 | -0,73 |
| C18 | 6,06 | -2,05 | -5,91 | 18,50 | -2,09 | -3,57 | -0,37 | -0,58 |
| C19 | -4,35 | 1,83 | -1,08 | 0,70 | -2,40 | 4,22 | -0,42 | -0,67 |
| C20 | -3,83 | -2,07 | 13,92 | -1,76 | -2,12 | -7,92 | -0,37 | -0,59 |
| C21 | -1,49 | -1,78 | 1,18 | -1,52 | 0,54 | 1,10 | -0,32 | -0,50 |
| C22 | 2,54 | -0,21 | -5,47 | -2,65 | -0,69 | 4,88 | -0,56 | -0,88 |
| C23 | -0,70 | 1,30 | 0,50 | -2,34 | 3,14 | -0,90 | -0,49 | -0,78 |
| C24 | -5,34 | 2,14 | 2,25 | -0,22 | -2,95 | 2,23 | -0,52 | -0,82 |
| C25 | 27,41 | -2,42 | -7,54 | -2,06 | -2,47 | -9,24 | -0,43 | -0,68 |
| C26 | -1,82 | -1,91 | -2,65 | 1,00 | 3,04 | 3,02 | -0,34 | -0,54 |

*Participations of agents in projects with specific technological domains (obs. freq)*

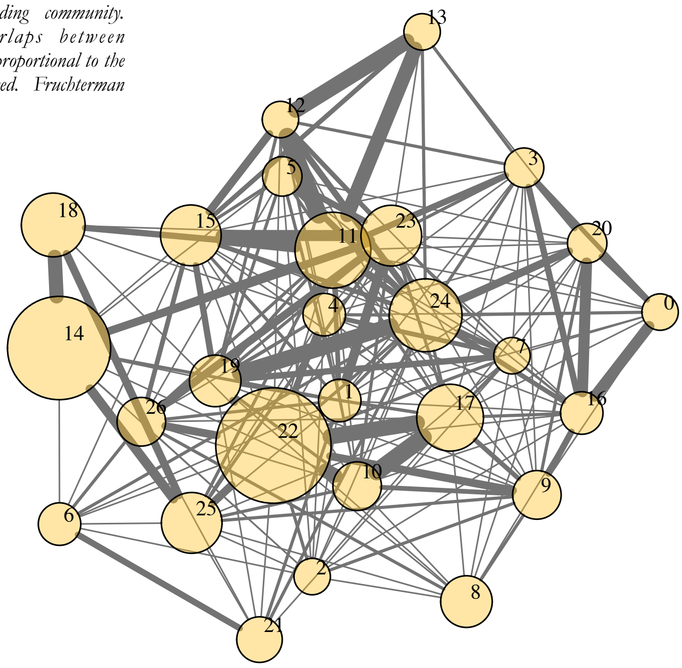
| CPM_k12 | Bio-Technologies | Organic Chemistry | Geotermal Sciences and Biomasses | ICT - Multimedia | Mechanics Engineering | Multi Disciplinary | Nanotechnologies | New Materials | Optoelectronic | NA |
|---|---|---|---|---|---|---|---|---|---|---|
| C0 | 17 | | | | | | 4 | | 57 | |
| C1 | 25 | | | 43 | | | 9 | 9 | 18 | |
| C2 | | | | 48 | | | | | | |
| C3 | 20 | | | | | | | | 57 | |
| C4 | 17 | | 8 | 45 | | | | | 7 | |
| C5 | 15 | | | 45 | | | 9 | | 8 | |
| C6 | | | | 54 | | | | | | |
| C7 | 34 | 5 | | 46 | | | | | | |
| C8 | 6 | | 4 | 47 | | | | | | |
| C9 | 10 | 10 | 7 | 46 | | | 8 | | 10 | 2 |
| C10 | | 5 | | 88 | 12 | | | | | |
| C11 | | 2 | | 70 | 15 | | 17 | | 32 | |
| C12 | | | | 45 | 4 | | 6 | | 12 | |
| C13 | | | | 20 | 4 | | 6 | | 15 | |
| C14 | 19 | 12 | | 89 | 36 | 5 | | | | |
| C15 | | 8 | | 63 | 5 | 5 | 15 | | 6 | |
| C16 | 17 | | | 8 | | | 2 | | 68 | |
| C17 | | | | 96 | 10 | | | 4 | 22 | |
| C18 | | 8 | | 35 | 40 | | | | 2 | |
| C19 | 17 | | 5 | 82 | | | | | 7 | |
| C20 | 29 | | | | | | | | 58 | |
| C21 | 25 | | | 36 | | | 4 | | | |
| C22 | | | 7 | 163 | | | 7 | 4 | 7 | |
| C23 | 25 | | | 67 | 22 | | 15 | | 20 | |
| C24 | 34 | | 5 | 84 | 4 | | | | 37 | |
| C25 | | | | 117 | | | | | | |
| C26 | | | 4 | 38 | 10 | | 16 | | 6 | |

*Participations of agents in projects with specific technological domains (adj. res.)*

| CPM_k12 | Bio-Technologies | Organic Chemistry | Geotermal Sciences and Biomasses | ICT - Multimedia | Mechanics Engineering | Multi Disciplinary | Nanotechnologies | New Materials | Optoelectronic | NA |
|---|---|---|---|---|---|---|---|---|---|---|
| C0 | 2,79 | -1,25 | -1,11 | -10,12 | -2,30 | -0,55 | 0,28 | -0,72 | 13,36 | -0,25 |
| C1 | 3,96 | -1,45 | -1,29 | -3,08 | -2,66 | -0,64 | 2,10 | 10,40 | 0,07 | -0,29 |
| C2 | -2,55 | -0,97 | -0,87 | 6,20 | -1,79 | -0,43 | -1,52 | -0,56 | -3,17 | -0,19 |
| C3 | 3,92 | -1,24 | -1,11 | -10,05 | -2,28 | -0,55 | -1,93 | -0,72 | 13,49 | -0,25 |
| C4 | 2,85 | -1,24 | 6,46 | 0,44 | -2,28 | -0,55 | -1,93 | -0,72 | -1,89 | -0,25 |
| C5 | 2,13 | -1,24 | -1,11 | 0,44 | -2,28 | -0,55 | 3,10 | -0,72 | -1,58 | -0,25 |
| C6 | -2,71 | -1,03 | -0,92 | 6,58 | -1,90 | -0,46 | -1,61 | -0,60 | -3,37 | -0,21 |
| C7 | 8,21 | 2,74 | -1,16 | -0,36 | -2,40 | -0,58 | -2,03 | -0,76 | -4,25 | -0,26 |
| C8 | -0,30 | -1,06 | 3,43 | 4,07 | -1,95 | -0,47 | -1,65 | -0,62 | -3,46 | -0,21 |
| C9 | -0,31 | 6,37 | 4,82 | -1,30 | -2,51 | -0,61 | 1,96 | -0,79 | -1,65 | 7,39 |
| C10 | -3,82 | 2,19 | -1,30 | 5,86 | 2,30 | -0,65 | -2,27 | -0,84 | -4,74 | -0,29 |
| C11 | -4,38 | -0,38 | -1,49 | -1,10 | 2,43 | -0,74 | 4,64 | -0,97 | 2,06 | -0,33 |
| C12 | -3,03 | -1,15 | -1,03 | 1,86 | -0,06 | -0,51 | 1,79 | -0,67 | 0,19 | -0,23 |
| C13 | -2,47 | -0,94 | -0,84 | -1,58 | 0,77 | -0,42 | 2,90 | -0,55 | 2,93 | -0,19 |
| C14 | 0,01 | 5,33 | -1,63 | -0,20 | 8,83 | 5,80 | -2,84 | -1,06 | -5,94 | -0,36 |
| C15 | -3,76 | 4,49 | -1,28 | 1,19 | -0,54 | 7,57 | 5,09 | -0,83 | -3,06 | -0,28 |
| C16 | 1,89 | -1,38 | -1,23 | -9,52 | -2,54 | -0,61 | -1,14 | -0,80 | 14,39 | -0,27 |
| C17 | -4,31 | -1,64 | -1,46 | 3,97 | 0,70 | -0,73 | -2,55 | 3,51 | -0,12 | -0,33 |
| C18 | -3,42 | 5,16 | -1,16 | -2,80 | 15,95 | -0,58 | -2,03 | -0,76 | -3,66 | -0,26 |
| C19 | 1,18 | -1,50 | 2,63 | 3,87 | -2,76 | -0,67 | -2,33 | -0,87 | -3,08 | -0,30 |
| C20 | 6,35 | -1,32 | -1,18 | -10,71 | -2,43 | -0,59 | -2,06 | -0,77 | 12,51 | -0,26 |
| C21 | 6,76 | -1,14 | -1,01 | -0,10 | -2,09 | -0,50 | 0,66 | -0,66 | -3,70 | -0,23 |
| C22 | -5,20 | -1,98 | 2,56 | 8,80 | -3,64 | -0,88 | -0,52 | 2,63 | -5,04 | -0,39 |
| C23 | 1,95 | -1,75 | -1,56 | -2,80 | 4,50 | -0,78 | 3,39 | -1,01 | -1,21 | -0,35 |
| C24 | 3,68 | -1,84 | 1,65 | -1,28 | -2,04 | -0,82 | -2,87 | -1,07 | 1,94 | -0,37 |
| C25 | -4,04 | -1,54 | -1,37 | 9,80 | -2,83 | -0,68 | -2,40 | -0,89 | -5,02 | -0,31 |
| C26 | -3,19 | -1,21 | 2,77 | -0,82 | 2,67 | -0,54 | 7,23 | -0,70 | -2,08 | -0,24 |

## CPM_k12

*Graph of the partition. Node are proportional to the number of agents that belong to the corresponding community. Edges represent overlaps between communities and they are proportional to the number of nodes involved. Fruchterman Reingold layout.*



## a.3. CPM_k18

*Typology of agents (obs. freq)*

| CPM_k18 | GI | TA | CC | SC | LGI | E | R | BS | KIBS |
|---|---|---|---|---|---|---|---|---|---|
| C0 | 1 | | 1 | | 5 | | 2 | | 10 |
| C1 | 1 | | | | 4 | 8 | 1 | | 6 |
| C2 | | | 2 | | 2 | 11 | | | 4 |
| C3 | 1 | 2 | | | 9 | 1 | | | 5 |
| C4 | 1 | 1 | | | 9 | 3 | | | 6 |
| C5 | | 1 | 2 | | 3 | 2 | | 1 | 11 |
| C6 | | 1 | 2 | 1 | 2 | 4 | 4 | 1 | 13 |
| C7 | | | 2 | | 5 | 2 | 4 | 1 | 4 |
| C8 | | | 1 | | 2 | 3 | 2 | | 12 |
| C9 | | | | 1 | | 2 | 1 | 1 | 15 |

*Typology of agents (adj. res)*

| CPM_k18 | GI | TA | CC | SC | LGI | E | R | BS | KIBS |
|---|---|---|---|---|---|---|---|---|---|
| C0 | 1,08 | -0,73 | 0,07 | -0,46 | 0,69 | -2,13 | 0,65 | -0,65 | 0,93 |
| C1 | 1,02 | -0,75 | -1,08 | -0,47 | -0,04 | 2,73 | -0,36 | -0,67 | -1,20 |
| C2 | -0,65 | -0,73 | 1,18 | -0,46 | -1,11 | 4,80 | -1,25 | -0,65 | -1,99 |
| C3 | 1,14 | 2,47 | -1,02 | -0,45 | 3,28 | -1,43 | -1,21 | -0,63 | -1,33 |
| C4 | 1,02 | 0,77 | -1,08 | -0,47 | 2,89 | -0,35 | -1,29 | -0,67 | -1,20 |
| C5 | -0,67 | 0,77 | 1,10 | -0,47 | -0,62 | -0,96 | -1,29 | 1,02 | 1,18 |
| C6 | -0,81 | 0,40 | 0,58 | 1,49 | -1,87 | -0,53 | 1,65 | 0,65 | 0,44 |
| C7 | -0,63 | -0,71 | 1,26 | -0,45 | 0,83 | -0,78 | 2,68 | 1,14 | -1,83 |
| C8 | -0,67 | -0,75 | 0,01 | -0,47 | -1,21 | -0,35 | 0,57 | -0,67 | 1,66 |
| C9 | -0,67 | -0,75 | -1,08 | 1,91 | -2,38 | -0,96 | -0,36 | 1,02 | 3,09 |

*Participations of agents in waves (obs. freq)*

| CPM_k18 | 2002_171 | 2002_172 | 2002_ITT | 2004_171 | 2004_171A | 2005_171 |
|---|---|---|---|---|---|---|
| C0 | | 7 | 8 | 8 | 15 | 46 |
| C1 | | 4 | 30 | | | 74 |
| C2 | 24 | | 2 | 29 | | 19 |
| C3 | | 8 | 7 | | 7 | 64 |
| C4 | 23 | 8 | 19 | | 7 | 49 |
| C5 | 12 | 5 | 30 | | 15 | 60 |
| C6 | 28 | 12 | 42 | 27 | | 31 |
| C7 | 11 | 10 | 32 | | 15 | 56 |
| C8 | | 8 | 65 | | | 74 |
| C9 | 117 | | | | | |

*Participations of agents in waves (adj. res.)*

| CPM_k18 | 2002_171 | 2002_172 | 2002_ITT | 2004_171 | 2004_171A | 2005_171 |
|---|---|---|---|---|---|---|
| C0 | -4,68 | 1,14 | -2,73 | 1,53 | 5,32 | 2,33 |
| C1 | -5,37 | -0,90 | 1,76 | -2,71 | -2,59 | 5,71 |
| C2 | 2,93 | -2,17 | -4,03 | 12,75 | -2,11 | -3,06 |
| C3 | -4,74 | 1,56 | -3,09 | -2,39 | 1,21 | 6,19 |
| C4 | 0,63 | 0,92 | -0,87 | -2,68 | 0,62 | 0,77 |
| C5 | -2,83 | -0,76 | 0,97 | -2,90 | 3,64 | 1,54 |
| C6 | 0,19 | 1,64 | 2,72 | 7,33 | -3,00 | -5,26 |
| C7 | -3,15 | 1,27 | 1,33 | -2,93 | 3,56 | 0,59 |
| C8 | -6,39 | -0,09 | 7,33 | -3,22 | -3,09 | 2,01 |
| C9 | 23,31 | -2,79 | -5,93 | -2,83 | -2,71 | -9,87 |

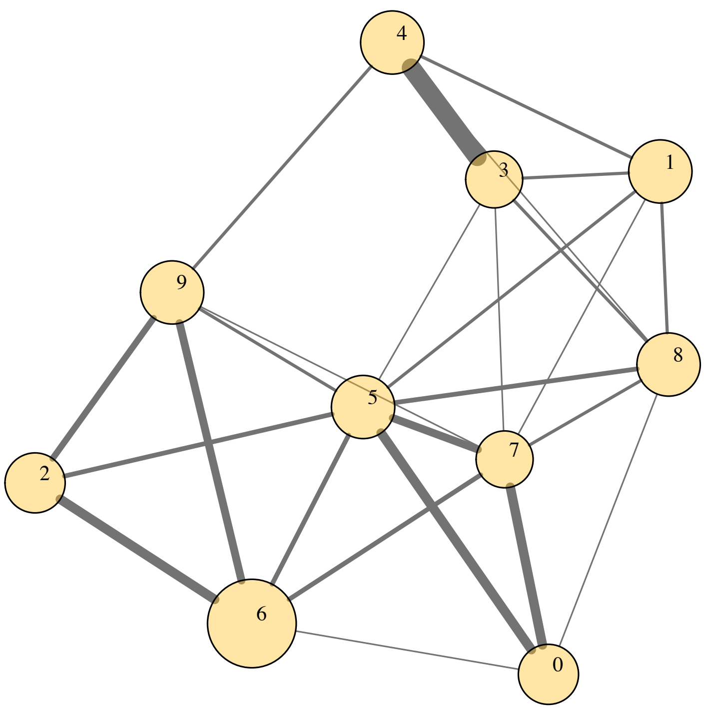
*Participations of agents in projects with specific technological domains (obs. freq)*

| CPM_k18 | Others | Bio-Technologies | Organic Chemistry | Geotermal Sciences and Biomasses | ICT - Multimedia | Mechanics Engineering | Multi Disciplinary | Nano technologies | New Materials | Opto electronic | NA |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **C0** | | 8 | | 45 | 5 | 5 | 15 | | 6 | 0 | 0 |
| **C1** | | | | 72 | 10 | | | 4 | 22 | 0 | 0 |
| **C2** | | 8 | | 28 | 36 | | | | 2 | 0 | 14 |
| **C3** | | | 7 | 68 | | | 7 | 4 | | 0 | 0 |
| **C4** | | | 7 | 92 | | | 7 | | | 0 | 0 |
| **C5** | | 2 | | 62 | 11 | | 17 | | 30 | 0 | 0 |
| **C6** | 19 | 12 | | 76 | 28 | 5 | | | | 0 | 0 |
| **C7** | 25 | | | 60 | 22 | | 15 | | 2 | 0 | 0 |
| **C8** | 34 | | | 72 | 4 | | | | 37 | 0 | 2 |
| **C9** | | | | 117 | | | | | | 2 | 0 |

*Participations of agents in projects with specific technological domains (adj. res.)*

| CPM_k18 | Bio-Technologies | Organic Chemistry | Geotermal Sciences and Biomasses | ICT - Multimedia | Mechanics Engineering | Multi Disciplinary | Nano technologies | New Materials | Opto electronic |
|---|---|---|---|---|---|---|---|---|---|
| **C0** | -2,62 | 4,00 | -1,08 | -1,75 | -1,41 | 5,09 | 5,16 | -0,81 | -0,60 |
| **C1** | -3,01 | -1,83 | -1,24 | 0,95 | -0,43 | -1,04 | -2,64 | 3,85 | 4,39 |
| **C2** | -2,45 | 4,45 | -1,01 | -4,53 | 11,11 | -0,85 | -2,15 | -0,76 | -1,95 |
| **C3** | -2,66 | -1,61 | 5,94 | 3,31 | -3,30 | -0,92 | 1,12 | 4,48 | -3,03 |
| **C4** | -2,98 | -1,81 | 5,18 | 5,44 | -3,70 | -1,03 | 0,52 | -0,92 | -3,39 |
| **C5** | -3,22 | -0,77 | -1,33 | -2,81 | -0,56 | -1,12 | 4,33 | -1,00 | 6,43 |
| **C6** | 3,23 | 4,57 | -1,43 | -2,14 | 3,94 | 3,57 | -3,06 | -1,08 | -3,97 |
| **C7** | 6,06 | -1,97 | -1,34 | -3,43 | 2,81 | -1,13 | 3,42 | -1,01 | -3,03 |
| **C8** | 8,19 | -2,17 | -1,47 | -3,62 | -3,30 | -1,24 | -3,14 | -1,11 | 7,41 |
| **C9** | -3,15 | -1,91 | -1,29 | 8,87 | -3,91 | -1,09 | -2,76 | -0,98 | -3,58 |

**CPM_k18**




*Graph of the partition. Node are proportional to the number of agents that belong to the corresponding community. Edges represent overlaps between communities and they are proportional to the number of nodes involved. Fruchterman Reingold layout.*

## a.4. MKV1

*Typology of agents (obs. freq)*

| MKV1 | GI | TA | CC | SC | LGI | E | R | BS | KIBS |
|---|---|---|---|---|---|---|---|---|---|
| C0 | 2 | 2 | 1 | 1 | 15 | 22 | 4 | | 24 |
| C1 | | 1 | 2 | 2 | 4 | 23 | 4 | 1 | 20 |
| C2 | 4 | 1 | 2 | | 7 | 3 | 2 | 1 | 15 |
| C3 | 3 | | | | 1 | 21 | 8 | | 14 |
| C4 | 1 | | 1 | | 1 | 9 | 3 | | 18 |
| C5 | 1 | 2 | 2 | 1 | 4 | 5 | 3 | | 16 |
| C6 | 3 | | 3 | | 4 | 12 | 2 | 1 | 3 |
| C7 | | 1 | 2 | | 3 | 3 | 1 | | 9 |
| C8 | | | 1 | 1 | 2 | 10 | 2 | | 6 |
| C9 | | | 1 | | 3 | 8 | 4 | | 7 |
| C10 | | | 1 | 1 | 1 | 4 | 1 | | 12 |
| C11 | 1 | | 4 | 1 | 2 | 5 | | 2 | 5 |
| C12 | 1 | 4 | | | | 4 | 1 | | 3 |
| C13 | | 1 | | | 2 | 8 | | | 4 |
| C14 | | | | | | 4 | | 1 | 4 |
| C15 | | | 2 | | 1 | 4 | 3 | 1 | 7 |
| C16 | | | 1 | | 2 | 6 | | 1 | 8 |
| C17 | 1 | | | | | 2 | 2 | 1 | 9 |
| C18 | | | | | | 1 | 1 | | 12 |
| C19 | | | | | | 2 | 2 | | 2 |
| C20 | | 2 | | | | 1 | | | 6 |
| C21 | | | | | | 2 | 3 | | 6 |
| C22 | | 1 | | | | 1 | | | |

*Typology of agents (adj. res)*

| MKV1 | GI | TA | CC | SC | LGI | E | R | BS | KIBS |
|---|---|---|---|---|---|---|---|---|---|
| C0 | -0,17 | 0,02 | -1,28 | 0,09 | 3,52 | 0,26 | -0,94 | -1,18 | -0,96 |
| C1 | -1,44 | -0,50 | -0,30 | 1,56 | -0,71 | 1,86 | -0,43 | 0,05 | -0,63 |
| C2 | 2,90 | 0,03 | 0,44 | -0,70 | 2,15 | -2,83 | -0,62 | 0,57 | 0,49 |
| C3 | 1,33 | -1,21 | -1,52 | -0,82 | -1,83 | 2,36 | 2,18 | -0,94 | -1,35 |
| C4 | -0,04 | -1,00 | -0,36 | -0,68 | -1,33 | -0,31 | 0,12 | -0,77 | 1,90 |
| C5 | -0,07 | 1,14 | 0,48 | 0,87 | 0,43 | -1,98 | 0,06 | -0,79 | 1,00 |
| C6 | 2,35 | -0,92 | 1,73 | -0,62 | 0,85 | 1,57 | -0,27 | 0,81 | -3,15 |
| C7 | -0,80 | 0,67 | 1,37 | -0,51 | 0,92 | -1,35 | -0,52 | -0,58 | 0,77 |
| C8 | -0,86 | -0,81 | 0,07 | 1,37 | -0,09 | 1,65 | 0,10 | -0,62 | -1,15 |
| C9 | -0,89 | -0,83 | 0,02 | -0,56 | 0,56 | 0,55 | 1,55 | -0,64 | -0,86 |
| C10 | -0,82 | -0,77 | 0,17 | 1,49 | -0,72 | -0,97 | -0,58 | -0,59 | 1,97 |
| C11 | 0,48 | -0,77 | 3,55 | 1,49 | 0,05 | -0,47 | -1,39 | 2,96 | -1,31 |
| C12 | 0,95 | 6,21 | -0,77 | -0,42 | -1,19 | 0,09 | -0,11 | -0,48 | -1,19 |
| C13 | -0,71 | 0,93 | -0,83 | -0,45 | 0,49 | 2,03 | -1,20 | -0,51 | -0,99 |
| C14 | -0,55 | -0,51 | -0,64 | -0,35 | -0,99 | 0,98 | -0,92 | 2,23 | 0,34 |
| C15 | -0,78 | -0,73 | 1,46 | -0,50 | -0,60 | -0,71 | 1,26 | 1,31 | -0,01 |
| C16 | -0,78 | -0,73 | 0,28 | -0,50 | 0,21 | 0,35 | -1,32 | 1,31 | 0,49 |
| C17 | 0,79 | -0,67 | -0,83 | -0,45 | -1,28 | -1,41 | 0,68 | 1,53 | 1,70 |
| C18 | -0,68 | -0,64 | -0,80 | -0,44 | -1,24 | -1,87 | -0,19 | -0,49 | 3,64 |
| C19 | -0,45 | -0,42 | -0,52 | -0,28 | -0,81 | 0,20 | 2,19 | -0,32 | -0,28 |
| C20 | -0,55 | 3,58 | -0,64 | -0,35 | -0,99 | -1,23 | -0,92 | -0,39 | 1,72 |
| C21 | -0,61 | -0,57 | -0,71 | -0,38 | -1,10 | -0,84 | 2,25 | -0,44 | 1,07 |
| C22 | -0,26 | 4,07 | -0,30 | -0,16 | -0,46 | 0,63 | -0,43 | -0,19 | -1,13 |

*Participations of agents in waves (obs. freq)*

| MKV1 | 2002 _171 | 2002 _172 | 2002 _ITT | 2004 _171 | 2004 _171 A | 2005 _171 | 2006 _VIN | 2007 _171 | 2008 _171 |
|---|---|---|---|---|---|---|---|---|---|
| C0 | 49 | 18 | 73 | 8 | 11 | 133 | 2 | | 2 |
| C1 | 47 | 14 | 51 | 37 | 2 | 50 | | 12 | |
| C2 | | 11 | 21 | 12 | 15 | 66 | | | |
| C3 | 6 | 2 | 146 | 8 | 2 | 10 | 4 | 8 | |
| C4 | | 22 | 19 | 9 | | 53 | 2 | 10 | |
| C5 | 20 | | 29 | 4 | 4 | 63 | | 5 | |
| C6 | 3 | 12 | 11 | 8 | | 43 | | 6 | |
| C7 | 8 | | 7 | | 8 | 38 | | 4 | |
| C8 | 21 | 10 | 15 | | 24 | 20 | | | |
| C9 | | 9 | 2 | 4 | 2 | 48 | | 6 | |
| C10 | 4 | 8 | 25 | | | 22 | 9 | 4 | |
| C11 | 4 | | 4 | 8 | 10 | 35 | | 7 | |
| C12 | | | 25 | | 2 | 2 | | | |
| C13 | | 10 | 54 | | | 11 | | 6 | |
| C14 | | | 4 | | | 18 | 2 | | |
| C15 | | 4 | 10 | 5 | | 35 | | 4 | |
| C16 | | | 40 | 12 | 23 | 3 | 5 | 6 | |
| C17 | | 15 | 52 | 16 | | | | 4 | |
| C18 | 4 | 21 | 7 | | | 17 | | 4 | |
| C19 | | | 31 | 4 | | | | | |
| C20 | | | 51 | | | 4 | | 10 | |
| C21 | | 3 | 74 | | | 3 | 6 | 2 | |
| C22 | | | | | 5 | | | | |

*Participations of agents in waves (adj. res.)*

| MKV1 | 2002 _171 | 2002 _172 | 2002 _ITT | 2004 _171 | 2004 _171 A | 2005 _171 | 2006 _VIN | 2007 _171 | 2008 _171 |
|---|---|---|---|---|---|---|---|---|---|
| C0 | 6,03 | -0,99 | -4,16 | -2,78 | -1,16 | 5,25 | -1,16 | -4,08 | 3,52 |
| C1 | 8,17 | -0,54 | -3,68 | 6,94 | -2,91 | -2,74 | -1,84 | 0,75 | -0,47 |
| C2 | -3,36 | 0,57 | -4,48 | 1,53 | 3,63 | 5,21 | -1,38 | -2,54 | -0,35 |
| C3 | -2,44 | -3,48 | 12,88 | -1,20 | -2,61 | -8,09 | 0,89 | -0,21 | -0,44 |
| C4 | -3,21 | 4,88 | -4,35 | 0,66 | -2,55 | 3,40 | 0,31 | 2,14 | -0,34 |
| C5 | 3,51 | -3,28 | -2,94 | -1,49 | -0,99 | 4,62 | -1,38 | -0,34 | -0,35 |
| C6 | -1,46 | 2,46 | -4,30 | 1,25 | -2,15 | 4,01 | -1,11 | 1,16 | -0,29 |
| C7 | 1,37 | -2,33 | -4,21 | -2,13 | 2,69 | 4,70 | -0,98 | 0,60 | -0,25 |
| C8 | 5,60 | 1,33 | -3,79 | -2,53 | 9,52 | -1,98 | -1,16 | -2,13 | -0,30 |
| C9 | -2,50 | 1,69 | -5,84 | -0,26 | -0,89 | 6,60 | -1,03 | 1,57 | -0,26 |
| C10 | -0,73 | 1,19 | -0,12 | -2,25 | -2,00 | -0,22 | 8,11 | 0,39 | -0,27 |
| C11 | -0,61 | -2,39 | -5,17 | 1,86 | 3,67 | 3,55 | -1,00 | 2,27 | -0,26 |
| C12 | -1,58 | -1,54 | 5,77 | -1,41 | 0,45 | -2,90 | -0,65 | -1,19 | -0,17 |
| C13 | -2,67 | 1,69 | 6,01 | -2,39 | -2,13 | -3,58 | -1,10 | 1,22 | -0,28 |
| C14 | -1,44 | -1,40 | -1,93 | -1,28 | -1,14 | 4,58 | 2,89 | -1,08 | -0,15 |
| C15 | -2,25 | -0,17 | -2,93 | 0,72 | -1,79 | 4,74 | -0,92 | 0,84 | -0,24 |
| C16 | -2,81 | -2,74 | 1,93 | 2,81 | 9,10 | -5,88 | 3,43 | 0,98 | -0,30 |
| C17 | -2,77 | 3,53 | 4,86 | 4,70 | -2,21 | -6,50 | -1,14 | -0,01 | -0,29 |
| C18 | -0,08 | 9,00 | -3,42 | -1,92 | -1,71 | 0,05 | -0,88 | 1,03 | -0,23 |
| C19 | -1,74 | -1,70 | 6,64 | 1,24 | -1,38 | -4,07 | -0,71 | -1,31 | -0,18 |
| C20 | -2,39 | -2,33 | 7,38 | -2,13 | -1,90 | -4,50 | -0,98 | 4,20 | -0,25 |
| C21 | -2,79 | -1,49 | 9,76 | -2,50 | -2,22 | -5,83 | 4,39 | -1,07 | -0,29 |
| C22 | -0,65 | -0,64 | -1,66 | -0,58 | 9,67 | -1,53 | -0,27 | -0,49 | -0,07 |

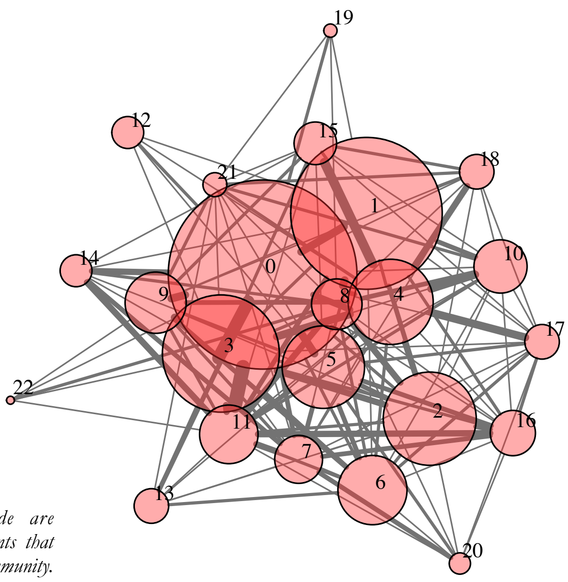
*Participations of agents in projects with specific technological domains (obs. freq)*

| MKV1 | Others | Bio-Technologies | Organic Chemistry | Geotermal Sciences and Biomasses | ICT - Multimedia | Mechanics Engineering | Multi Disciplinary | Nanotechnologies | New Materials | Optoelectronic | NA |
|---|---|---|---|---|---|---|---|---|---|---|---|
| C0 | 2 | | 11 | 13 | 202 | 12 | | 12 | 8 | 30 | 6 |
| C1 | | 19 | 14 | | 110 | 61 | 5 | | 2 | 2 | |
| C2 | | | 12 | 4 | 55 | 18 | | 17 | | 19 | |
| C3 | | 28 | 4 | 4 | 8 | | | 6 | | 132 | 4 |
| C4 | | 6 | 2 | 9 | 70 | 4 | | | | 24 | |
| C5 | | 25 | | 4 | 84 | 6 | | 4 | | | 2 |
| C6 | | 8 | | 8 | 43 | 18 | | | | 2 | 4 |
| C7 | | | | 8 | 38 | 6 | | | | 9 | 4 |
| C8 | | 10 | 6 | | 41 | 11 | | 9 | 13 | | |
| C9 | | 3 | 13 | | 39 | | | | | 8 | 8 |
| C10 | | 3 | | | 58 | 3 | | | | 8 | |
| C11 | | | 21 | | 24 | 2 | 3 | 2 | 14 | 2 | |
| C12 | | 21 | 2 | 2 | 4 | | | | | | |
| C13 | 18 | | | | 5 | | | 2 | | 56 | |
| C14 | | | | | 21 | | 3 | | | | |
| C15 | | 9 | | 5 | 18 | | 8 | | | 18 | |
| C16 | | 34 | 24 | 19 | 5 | 3 | | 4 | | | |
| C17 | | 38 | | 16 | 19 | | | | | 14 | |
| C18 | | 20 | | | 26 | | | | | 7 | |
| C19 | | | 4 | | 28 | | | 3 | | | |
| C20 | | 51 | | | 4 | | | | | | 10 |
| C21 | | 6 | | | 18 | | | | | 62 | 2 |
| C22 | | | 5 | | | | | | | | |

*Participations of agents in projects with specific technological domains (adj. res.)*

| MKV1 | Others | Bio-Technologies | Organic Chemistry | Geotermal Sciences and Biomasses | ICT - Multimedia | Mechanics Engineering | Multi Disciplinary | Nanotechnologies | New Materials | Optoelectronic | NA |
|---|---|---|---|---|---|---|---|---|---|---|---|
| C0 | -0,51 | -7,24 | -1,49 | 0,05 | 9,32 | -2,01 | -1,76 | 1,44 | 1,36 | -4,00 | 0,20 |
| C1 | -1,50 | -1,96 | 0,68 | -3,28 | 2,58 | 13,37 | 2,37 | -2,60 | -0,95 | -6,96 | -2,13 |
| C2 | -1,12 | -4,50 | 2,03 | -0,64 | 0,16 | 3,49 | -1,10 | 7,59 | -1,54 | -0,98 | -1,60 |
| C3 | -1,39 | 0,77 | -2,12 | -1,53 | -11,25 | -3,85 | -1,36 | 0,39 | -1,90 | 19,28 | 0,28 |
| C4 | -1,08 | -2,61 | -1,84 | 1,89 | 3,90 | -1,45 | -1,05 | -1,86 | -1,47 | 0,67 | -1,53 |
| C5 | -1,12 | 2,30 | -2,80 | -0,64 | 5,55 | -0,91 | -1,10 | 0,30 | -1,54 | -5,49 | -0,24 |
| C6 | -0,91 | -0,99 | -2,26 | 2,42 | 1,59 | 5,51 | -0,88 | -1,57 | -1,24 | -3,85 | 2,01 |
| C7 | -0,80 | -3,20 | -1,99 | 3,21 | 2,50 | 0,80 | -0,78 | -1,38 | -1,09 | -0,98 | 2,57 |
| C8 | -0,95 | -0,61 | 0,47 | -2,06 | 0,43 | 2,10 | -0,92 | 4,26 | 9,41 | -4,62 | -1,34 |
| C9 | -0,84 | -2,28 | 4,77 | -1,82 | 2,01 | -2,31 | -0,81 | -1,45 | -1,14 | -1,60 | 5,92 |
| C10 | -0,84 | -2,31 | -2,09 | -1,84 | 6,48 | -0,90 | -0,82 | -1,46 | -1,15 | -1,65 | -1,20 |
| C11 | -0,82 | -3,27 | 9,26 | -1,78 | -1,36 | -1,28 | 3,13 | 0,08 | 12,07 | -3,36 | -1,16 |
| C12 | -0,53 | 9,47 | 0,32 | 0,68 | -3,23 | -1,46 | -0,52 | -0,92 | -0,72 | -2,58 | -0,75 |
| C13 | 20,22 | -3,58 | -2,23 | -1,95 | -6,88 | -2,48 | -0,87 | -0,17 | -1,22 | 11,96 | -1,27 |
| C14 | -0,48 | -1,92 | -1,20 | -1,05 | 4,39 | -1,33 | 6,07 | -0,83 | -0,66 | -2,35 | -0,68 |
| C15 | -0,75 | 0,52 | -1,87 | 1,63 | -1,92 | -2,08 | 10,58 | -1,31 | -1,03 | 2,49 | -1,07 |
| C16 | -0,94 | 7,10 | 9,01 | 8,05 | -7,34 | -1,31 | -0,92 | 1,01 | -1,28 | -4,59 | -1,34 |
| C17 | -0,93 | 8,56 | -2,31 | 6,58 | -4,13 | -2,57 | -0,91 | -1,61 | -1,27 | -0,59 | -1,32 |
| C18 | -0,72 | 5,33 | -1,79 | -1,57 | 0,85 | -1,99 | -0,70 | -1,25 | -0,98 | -1,01 | -1,02 |
| C19 | -0,58 | -2,33 | 1,53 | -1,27 | 4,41 | -1,61 | -0,57 | 2,10 | -0,79 | -2,84 | -0,83 |
| C20 | -0,80 | 15,76 | -1,99 | -1,74 | -6,14 | -2,21 | -0,78 | -1,38 | -1,09 | -3,90 | 8,13 |
| C21 | -0,93 | -1,82 | -2,32 | -2,04 | -4,42 | -2,59 | -0,91 | -1,62 | -1,28 | 12,81 | 0,27 |
| C22 | -0,22 | -0,87 | 9,23 | -0,48 | -1,96 | -0,60 | -0,21 | -0,38 | -0,30 | -1,07 | -0,31 |

## Mkv1




*Graph of the partition. Node are proportional to the number of agents that belong to the corresponding community. Edges represent overlaps between communities and they are proportional to the number of nodes involved. Fruchterman Reingold layout.*

## a.5. MKV2_AS

*Typology of agents (obs. freq)*

| MKV2_AS | GI | TA | CC | SC | LGI | E | R | BS | KIBS |
|---|---|---|---|---|---|---|---|---|---|
| C0 | | 1 | 2 | 1 | 2 | 14 | 4 | 1 | 15 |
| C1 | 5 | 1 | 3 | | 8 | 5 | 8 | 2 | 29 |
| C2 | 1 | 2 | | | 11 | 7 | 1 | | 14 |
| C3 | | | 1 | | 4 | 6 | 4 | | 22 |
| C4 | 3 | 1 | | | 6 | 20 | 8 | | 19 |
| C5 | 1 | | 2 | | 7 | 9 | 3 | | 11 |
| C6 | 1 | 3 | 2 | | 5 | 6 | 3 | | 14 |
| C7 | | | 2 | | 2 | 12 | 2 | | 7 |
| C8 | | | | 1 | | 2 | 1 | 1 | 15 |
| C9 | 3 | | | | 5 | 6 | 1 | 1 | 1 |
| C10 | | | 1 | | 2 | 6 | 3 | | 5 |
| C11 | | | 1 | | 4 | 3 | 2 | | 9 |
| C12 | | | 1 | | 2 | 3 | | | 10 |
| C13 | | | 1 | | 3 | 2 | 2 | | 9 |
| C14 | | | 10 | | | | | | 5 |
| C15 | | | 1 | | 1 | 2 | 4 | | 5 |
| C16 | | | | | 2 | 4 | 1 | 1 | 7 |
| C17 | | | | 1 | | 3 | 1 | | 7 |
| C18 | 1 | | 2 | | 1 | 1 | 2 | 2 | 4 |
| C19 | | | 1 | | 1 | 3 | 4 | | 4 |
| C20 | | | | | 1 | 4 | 3 | | 6 |
| C21 | | | | | 3 | 6 | 2 | | 4 |
| C22 | 1 | 2 | | | | 5 | 3 | | 3 |
| C23 | | | | | 2 | | 3 | 1 | 4 |
| C24 | | | | | | 5 | | | 7 |
| C25 | | | | | | | 2 | | 10 |
| C26 | | 1 | 1 | | | | 1 | | 6 |
| C27 | | | | | | 1 | 1 | | 3 |
| C28 | 1 | | | | 1 | 3 | 1 | 1 | 4 |
| C29 | | | | 1 | | | 2 | | 5 |
| C30 | 2 | | | | 1 | 2 | | | 2 |
| C31 | | | | | 2 | 2 | 1 | | 2 |
| C32 | | | 1 | | 1 | 1 | 1 | | 4 |
| C33 | 1 | | | | | 3 | 1 | | 4 |
| C34 | | | 1 | | | 4 | 2 | | 6 |
| C35 | | 1 | | | | | | | 6 |
| C36 | | | | | | 2 | 1 | | 3 |
| C37 | 1 | | | | 1 | 3 | | | 2 |
| C38 | | | | | | 1 | 2 | | 5 |
| C39 | | | | 1 | | 3 | | | 4 |
| C40 | | | | | 1 | 1 | 1 | | 4 |
| C41 | | | | | | 1 | 1 | | 5 |
| C42 | | | | | | 3 | | 1 | 2 |
| C43 | | 1 | 1 | | | 1 | | | 2 |
| C44 | | | | | | 1 | | | 3 |
| C45 | | | | | | 1 | 1 | 1 | 1 |
| C46 | | | | | | | 4 | | 1 |
| C47 | | | | | 2 | | | | 2 |
| C48 | | | | | | 1 | 3 | | 1 |
| C49 | | | | | 1 | | 1 | | 2 |
| C50 | | 1 | | | | 1 | | | 1 |
| C51 | | | | | | | | | 2 |
| C52 | | | | | | 1 | | | 2 |
| C53 | | | | | | 1 | 2 | | 1 |

*Typology of agents (adj. res)*

| MKV2_AS | GI | TA | CC | SC | LGI | E | R | BS | KIBS |
|---|---|---|---|---|---|---|---|---|---|
| C0 | -1,09 | 0,32 | 0,17 | 1,49 | -1,21 | 1,96 | -0,44 | 0,48 | -0,77 |
| C1 | 2,71 | -0,12 | 0,18 | -0,66 | 0,62 | -2,78 | 0,23 | 1,12 | 0,68 |
| C2 | 0,01 | 1,70 | -1,33 | -0,50 | 3,93 | -0,44 | -1,77 | -0,78 | -0,56 |
| C3 | -1,05 | -0,85 | -0,53 | -0,51 | 0,01 | -0,93 | -0,26 | -0,79 | 2,02 |
| C4 | 1,21 | -0,05 | -1,70 | -0,64 | -0,06 | 2,39 | 0,44 | -0,99 | -1,59 |
| C5 | 0,10 | -0,80 | 0,46 | -0,48 | 1,98 | 0,69 | -0,56 | -0,74 | -1,19 |
| C6 | 0,07 | 3,11 | 0,41 | -0,48 | 0,76 | -0,68 | -0,61 | -0,75 | -0,27 |
| C7 | -0,86 | -0,70 | 0,87 | -0,41 | -0,45 | 3,12 | -0,65 | -0,64 | -1,58 |
| C8 | -0,76 | -0,62 | -0,98 | 2,44 | -1,57 | -1,35 | -1,00 | 1,25 | 2,89 |
| C9 | 3,80 | -0,57 | -0,90 | -0,34 | 2,51 | 1,29 | -0,80 | 1,44 | -3,16 |
| C10 | -0,70 | -0,57 | 0,29 | -0,34 | 0,14 | 1,29 | 0,70 | -0,53 | -1,18 |
| C11 | -0,74 | -0,60 | 0,17 | -0,36 | 1,47 | -0,70 | -0,22 | -0,56 | 0,36 |
| C12 | -0,68 | -0,55 | 0,35 | -0,33 | 0,23 | -0,36 | -1,51 | -0,51 | 1,56 |
| C13 | -0,70 | -0,57 | 0,29 | -0,34 | 0,93 | -1,07 | -0,05 | -0,53 | 0,80 |
| C14 | -0,66 | -0,54 | 11,79 | -0,32 | -1,36 | -2,10 | -1,46 | -0,49 | -0,79 |
| C15 | -0,61 | -0,50 | 0,57 | -0,30 | -0,36 | -0,61 | 2,07 | -0,46 | -0,36 |
| C16 | -0,66 | -0,54 | -0,85 | -0,32 | 0,33 | 0,40 | -0,66 | 1,60 | 0,26 |
| C17 | -0,59 | -0,48 | -0,75 | 3,32 | -1,21 | 0,22 | -0,41 | -0,44 | 1,05 |
| C18 | 1,10 | -0,50 | 1,93 | -0,30 | -0,36 | -1,28 | 0,36 | 4,04 | -0,93 |
| C19 | -0,61 | -0,50 | 0,57 | -0,30 | -0,36 | 0,06 | 2,07 | -0,46 | -0,93 |
| C20 | -0,64 | -0,52 | -0,82 | -0,31 | -0,44 | 0,56 | 1,07 | -0,48 | -0,04 |
| C21 | -0,66 | -0,54 | -0,85 | -0,32 | 1,17 | 1,65 | 0,14 | -0,49 | -1,32 |
| C22 | 1,01 | 3,50 | -0,82 | -0,31 | -1,31 | 1,21 | 1,07 | -0,48 | -1,67 |
| C23 | -0,54 | -0,44 | -0,69 | -0,26 | 0,95 | -1,71 | 1,73 | 2,16 | -0,22 |
| C24 | -0,59 | -0,48 | -0,75 | -0,28 | -1,21 | 1,61 | -1,30 | -0,44 | 1,05 |
| C25 | -0,59 | -0,48 | -0,75 | -0,28 | -1,21 | -1,88 | 0,48 | -0,44 | 2,82 |
| C26 | -0,51 | 2,09 | 0,97 | -0,25 | -1,05 | -1,62 | -0,10 | -0,38 | 1,42 |
| C27 | -0,38 | -0,31 | -0,48 | -0,18 | -0,78 | -0,13 | 0,54 | -0,28 | 0,75 |
| C28 | 1,29 | -0,46 | -0,72 | -0,27 | -0,18 | 0,39 | -0,32 | 2,02 | -0,47 |
| C29 | -0,48 | -0,39 | -0,61 | 4,17 | -0,99 | -1,53 | 1,11 | -0,36 | 1,10 |
| C30 | 4,20 | -0,36 | -0,57 | -0,22 | 0,30 | 0,39 | -0,99 | -0,34 | -0,79 |
| C31 | -0,45 | -0,36 | -0,57 | -0,22 | 1,53 | 0,39 | 0,17 | -0,34 | -0,79 |
| C32 | -0,48 | -0,39 | 1,11 | -0,23 | 0,16 | -0,68 | 0,03 | -0,36 | 0,38 |
| C33 | 1,54 | -0,41 | -0,65 | -0,25 | -1,05 | 0,79 | -0,10 | -0,38 | 0,07 |
| C34 | -0,61 | -0,50 | 0,57 | -0,30 | -1,26 | 0,73 | 0,36 | -0,46 | 0,20 |
| C35 | -0,45 | 2,47 | -0,57 | -0,22 | -0,92 | -1,43 | -0,99 | -0,34 | 2,27 |
| C36 | -0,41 | -0,34 | -0,53 | -0,20 | -0,85 | 0,64 | 0,34 | -0,31 | 0,33 |
| C37 | 1,87 | -0,36 | -0,57 | -0,22 | 0,30 | 1,30 | -0,99 | -0,34 | -0,79 |
| C38 | -0,48 | -0,39 | -0,61 | -0,23 | -0,99 | -0,68 | 1,11 | -0,36 | 1,10 |
| C39 | -0,48 | -0,39 | -0,61 | 4,17 | -0,99 | 1,03 | -1,06 | -0,36 | 0,38 |
| C40 | -0,45 | -0,36 | -0,57 | -0,22 | 0,30 | -0,52 | 0,17 | -0,34 | 0,74 |
| C41 | -0,45 | -0,36 | -0,57 | -0,22 | -0,92 | -0,52 | 0,17 | -0,34 | 1,50 |
| C42 | -0,41 | -0,34 | -0,53 | -0,20 | -0,85 | 1,63 | -0,92 | 2,98 | -0,50 |
| C43 | -0,38 | 3,04 | 1,69 | -0,18 | -0,78 | -0,13 | -0,84 | -0,28 | -0,15 |
| C44 | -0,34 | -0,27 | -0,43 | -0,16 | -0,70 | 0,12 | -0,75 | -0,25 | 1,28 |
| C45 | -0,34 | -0,27 | -0,43 | -0,16 | -0,70 | 0,12 | 0,79 | 3,78 | -0,74 |
| C46 | -0,38 | -0,31 | -0,48 | -0,18 | -0,78 | -1,21 | 4,65 | -0,28 | -1,06 |
| C47 | -0,34 | -0,27 | -0,43 | -0,16 | 2,54 | -1,08 | -0,75 | -0,25 | 0,27 |
| C48 | -0,38 | -0,31 | -0,48 | -0,18 | -0,78 | -0,13 | 3,28 | -0,28 | -1,06 |
| C49 | -0,34 | -0,27 | -0,43 | -0,16 | 0,92 | -1,08 | 0,79 | -0,25 | 0,27 |
| C50 | -0,29 | 4,07 | -0,38 | -0,14 | -0,60 | 0,46 | -0,65 | -0,22 | -0,35 |
| C51 | -0,24 | -0,19 | -0,31 | -0,12 | -0,49 | -0,76 | -0,53 | -0,18 | 1,62 |
| C52 | -0,29 | -0,24 | -0,38 | -0,14 | -0,60 | 0,46 | -0,65 | -0,22 | 0,82 |
| C53 | -0,34 | -0,27 | -0,43 | -0,16 | -0,70 | 0,12 | 2,32 | -0,25 | -0,74 |

*Participations of agents in waves (obs. freq)*

| MKV2_AS | 2002_ 171 | 2002_ 172 | 2002_ ITT | 2004_ 171 | 2004_ 171A | 2005_ 171 | 2006_ VIN | 2007_ 171 | 2008_ 171 |
|---|---|---|---|---|---|---|---|---|---|
| C0 | 39 | 12 | 46 | 27 | 2 | 36 | 3 | 10 | |
| C1 | 28 | 19 | 82 | 12 | 15 | 143 | | 2 | |
| C2 | 38 | 8 | 26 | | 7 | 104 | | | |
| C3 | | 27 | 82 | 5 | | 103 | | 6 | |
| C4 | 22 | 14 | 173 | 2 | 17 | 34 | 4 | 8 | |
| C5 | 30 | 6 | 46 | | | 102 | | | |
| C6 | 13 | 15 | 46 | | 4 | 71 | | 3 | |
| C7 | 35 | | 20 | 33 | | 23 | | | |
| C8 | 117 | | | | | | | | |
| C9 | 3 | 6 | 6 | 4 | | 38 | | | |
| C10 | 11 | 10 | 45 | | 22 | 24 | | | |
| C11 | | 5 | 36 | | | 66 | | | |
| C12 | 8 | | 41 | | 10 | 37 | | | |
| C13 | | 8 | 35 | | 9 | 28 | 7 | | |
| C14 | 8 | | 41 | 8 | 4 | 40 | | | |
| C15 | 8 | 5 | 35 | | | 29 | | 4 | |
| C16 | 3 | 8 | 4 | | | 40 | 2 | | |
| C17 | 33 | | 23 | | | 10 | | | 2 |
| C18 | | 10 | 11 | 8 | 8 | 48 | | | |
| C19 | | | 39 | 5 | | 27 | | | |
| C20 | 11 | 20 | 15 | | | 22 | | 2 | |
| C21 | 24 | 18 | 12 | | 7 | 7 | | 6 | |
| C22 | 23 | | 43 | | | | | | |
| C23 | | 4 | 18 | | 7 | 28 | | | |
| C24 | 34 | | 12 | 4 | | 7 | 2 | | |
| C25 | 8 | 19 | 16 | | | 10 | | 2 | |
| C26 | 5 | | 17 | | 4 | 9 | | 6 | |
| C27 | | 8 | | | | 11 | | | |
| C28 | | | 24 | 16 | 2 | | 3 | 5 | |
| C29 | 15 | | 33 | | | | | | |
| C30 | | 8 | 8 | | | 8 | 2 | 4 | |
| C31 | | 2 | 24 | | | 9 | | 2 | |
| C32 | | | 17 | | | 28 | | | |
| C33 | 19 | 8 | 8 | 6 | | 5 | | | 2 |
| C34 | 35 | | 32 | 12 | | | | | |
| C35 | 8 | | 21 | | | 31 | | | |
| C36 | | 3 | 8 | | | 3 | 6 | | |
| C37 | | 22 | 6 | | | | | | |
| C38 | | | 54 | | | | | | |
| C39 | 23 | | 2 | 8 | 8 | 2 | 3 | 3 | |
| C40 | | 5 | 3 | | | 12 | | 4 | |
| C41 | 4 | 17 | | | | | | 4 | |
| C42 | | | | | 13 | 3 | 5 | | |
| C43 | | | | | | 6 | | 8 | |
| C44 | | | 17 | | | 4 | | 4 | |
| C45 | | 3 | | 3 | | | | 4 | |
| C46 | | | 39 | | | | 2 | 2 | |
| C47 | | | 13 | | | 2 | | 2 | |
| C48 | 19 | 5 | 4 | 2 | | | | | |
| C49 | | 3 | 4 | | 9 | | | | |
| C50 | | 3 | | | 5 | | | | |
| C51 | | 3 | | | | | | 2 | |
| C52 | 8 | | 2 | | 8 | | | | |
| C53 | | 11 | | | | | | | |

*Participations of agents in waves (adj. res.)*

| MKV2_AS | 2002_ 171 | 2002_ 172 | 2002_ ITT | 2004_ 171 | 2004_ 171A | 2005_ 171 | 2006_ VIN | 2007_ 171 | 2008_ 171 |
|---|---|---|---|---|---|---|---|---|---|
| C0 | 2,23 | -0,61 | -1,95 | 7,93 | -2,03 | -3,06 | 0,97 | 2,95 | -0,43 |
| C1 | -3,39 | -1,17 | -2,24 | 0,01 | 0,77 | 6,43 | -1,82 | -2,04 | -0,58 |
| C2 | 1,71 | -1,89 | -5,56 | -2,82 | -0,21 | 7,72 | -1,39 | -2,17 | -0,44 |
| C3 | -6,77 | 2,27 | 1,21 | -1,37 | -3,19 | 5,04 | -1,55 | 0,31 | -0,49 |
| C4 | -3,81 | -1,87 | 10,97 | -2,85 | 1,79 | -6,91 | 0,79 | 0,60 | -0,55 |
| C5 | 0,03 | -2,46 | -2,38 | -2,83 | -2,88 | 7,33 | -1,40 | -2,17 | -0,45 |
| C6 | -2,61 | 0,83 | -0,75 | -2,56 | -0,95 | 4,26 | -1,26 | -0,34 | -0,40 |
| C7 | 4,44 | -3,17 | -3,42 | 14,09 | -2,22 | -2,38 | -1,08 | -1,67 | -0,34 |
| C8 | 24,97 | -3,26 | -7,72 | -2,24 | -2,28 | -7,37 | -1,10 | -1,72 | -0,35 |
| C9 | -2,26 | 0,68 | -3,64 | 1,18 | -1,58 | 5,86 | -0,77 | -1,19 | -0,24 |
| C10 | -1,86 | 0,33 | 1,62 | -2,19 | 8,37 | -2,23 | -1,08 | -1,68 | -0,34 |
| C11 | -4,61 | -1,31 | 0,13 | -2,13 | -2,18 | 6,95 | -1,05 | -1,64 | -0,34 |
| C12 | -2,12 | -2,94 | 2,03 | -2,02 | 3,13 | 1,61 | -1,00 | -1,55 | -0,32 |
| C13 | -4,15 | 0,39 | 1,44 | -1,92 | 2,95 | 0,23 | 6,68 | -1,48 | -0,30 |
| C14 | -2,29 | -3,02 | 1,63 | 2,06 | -0,09 | 1,88 | -1,02 | -1,59 | -0,33 |
| C15 | -1,56 | -0,64 | 1,96 | -1,85 | -1,89 | 0,94 | -0,91 | 1,52 | -0,29 |
| C16 | -2,26 | 1,66 | -4,21 | -1,55 | -1,58 | 6,43 | 1,92 | -1,19 | -0,24 |
| C17 | 7,29 | -2,47 | 0,13 | -1,69 | -1,73 | -2,94 | -0,84 | -1,30 | 7,38 |
| C18 | -4,10 | 1,26 | -3,99 | 2,59 | 2,47 | 5,12 | -0,94 | -1,46 | -0,30 |
| C19 | -3,74 | -2,52 | 3,95 | 1,33 | -1,77 | 1,28 | -0,86 | -1,33 | -0,27 |
| C20 | -0,11 | 6,35 | -2,09 | -1,72 | -1,75 | 0,07 | -0,85 | 0,26 | -0,27 |
| C21 | 3,82 | 5,18 | -3,11 | -1,77 | 2,33 | -4,05 | -0,87 | 3,26 | -0,28 |
| C22 | 4,14 | -2,43 | 5,59 | -1,67 | -1,70 | -5,50 | -0,82 | -1,28 | -0,26 |
| C23 | -3,35 | -0,30 | -0,24 | -1,55 | 3,12 | 2,97 | -0,77 | -1,19 | -0,24 |
| C24 | 8,70 | -2,30 | -2,09 | 1,11 | -1,61 | -3,21 | 1,86 | -1,21 | -0,25 |
| C25 | -0,34 | 7,25 | -0,63 | -1,52 | -1,55 | -2,08 | -0,75 | 0,61 | -0,24 |
| C26 | -0,70 | -1,91 | 1,15 | -1,31 | 1,82 | -1,27 | -0,65 | 5,17 | -0,21 |
| C27 | -1,92 | 5,46 | -3,07 | -0,89 | -0,91 | 2,54 | -0,44 | -0,68 | -0,14 |
| C28 | -3,13 | -2,11 | 2,26 | 10,21 | -0,05 | -4,78 | 3,58 | 3,55 | -0,23 |
| C29 | 2,84 | -2,07 | 5,29 | -1,42 | -1,45 | -4,68 | -0,70 | -1,09 | -0,22 |
| C30 | -2,42 | 3,75 | -0,75 | -1,12 | -1,14 | -0,52 | 3,13 | 3,94 | -0,18 |
| C31 | -2,69 | -0,60 | 4,13 | -1,24 | -1,27 | -0,89 | -0,61 | 1,21 | -0,20 |
| C32 | -2,97 | -2,00 | 0,68 | -1,37 | -1,40 | 4,55 | -0,68 | -1,06 | -0,22 |
| C33 | 4,42 | 2,20 | -2,43 | 3,04 | -1,45 | -3,11 | -0,70 | -1,09 | 8,85 |
| C34 | 6,84 | -2,66 | 1,42 | 5,15 | -1,86 | -6,03 | -0,90 | -1,40 | -0,29 |
| C35 | -0,61 | -2,32 | 0,32 | -1,59 | -1,62 | 3,48 | -0,79 | -1,22 | -0,25 |
| C36 | -1,97 | 1,14 | 0,66 | -0,91 | -0,93 | -1,56 | 13,07 | -0,70 | -0,14 |
| C37 | -2,34 | 13,74 | -1,31 | -1,08 | -1,10 | -3,56 | -0,53 | -0,83 | -0,17 |
| C38 | -3,26 | -2,19 | 10,53 | -1,51 | -1,54 | -4,97 | -0,74 | -1,16 | -0,24 |
| C39 | 5,87 | -2,09 | -4,34 | 4,45 | 4,32 | -4,11 | 3,63 | 1,73 | -0,23 |
| C40 | -2,16 | 2,30 | -2,15 | -1,00 | -1,02 | 2,01 | -0,49 | 4,60 | -0,16 |
| C41 | -0,03 | 11,03 | -3,53 | -1,02 | -1,04 | -3,37 | -0,50 | 4,48 | -0,16 |
| C42 | -2,02 | -1,36 | -3,23 | -0,94 | 13,34 | -1,66 | 10,53 | -0,72 | -0,15 |
| C43 | -1,65 | -1,11 | -2,64 | -0,76 | -0,78 | 0,96 | -0,38 | 13,45 | -0,12 |
| C44 | -2,21 | -1,49 | 3,72 | -1,02 | -1,04 | -1,63 | -0,50 | 4,48 | -0,16 |
| C45 | -1,39 | 2,55 | -2,23 | 4,22 | -0,66 | -2,12 | -0,32 | 7,80 | -0,10 |
| C46 | -2,90 | -1,96 | 8,08 | -1,34 | -1,37 | -4,42 | 2,42 | 0,98 | -0,21 |
| C47 | -1,82 | -1,23 | 3,81 | -0,84 | -0,86 | -1,72 | -0,42 | 2,54 | -0,13 |
| C48 | 7,03 | 1,73 | -2,31 | 0,76 | -1,14 | -3,69 | -0,55 | -0,86 | -0,18 |
| C49 | -1,76 | 1,57 | -0,69 | -0,82 | 10,50 | -2,69 | -0,40 | -0,63 | -0,13 |
| C50 | -1,25 | 3,06 | -1,99 | -0,58 | 8,31 | -1,90 | -0,29 | -0,44 | -0,09 |
| C51 | -0,98 | 4,26 | -1,57 | -0,46 | -0,46 | -1,50 | -0,23 | 5,52 | -0,07 |
| C52 | 3,26 | -1,26 | -1,99 | -0,87 | 8,62 | -2,85 | -0,43 | -0,67 | -0,14 |
| C53 | -1,46 | 11,20 | -2,34 | -0,68 | -0,69 | -2,23 | -0,33 | -0,52 | -0,11 |

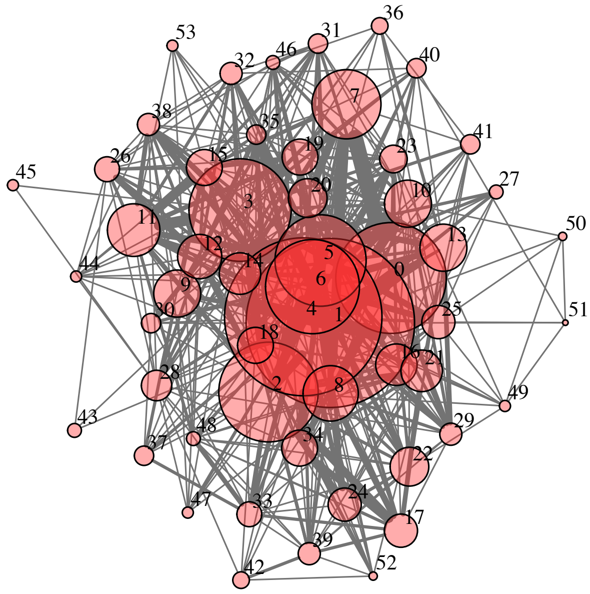
*Participations of agents in projects with specific technological domains (obs. freq)*

| MKV2_AS | Others | Bio-Technologies | Organic Chemistry | Geotermal Sciences and Biomasses | ICT - Multimedia | Mechanics Engineering | Multi Disciplinary | Nanotechnologies | New Materials | Optoelectronic | NA |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **C0** | | 19 | 12 | | 98 | 39 | 5 | | 2 | | |
| **C1** | | 25 | 10 | 4 | 153 | 37 | 5 | 24 | | 43 | |
| **C2** | | | | 7 | 158 | | | 7 | 4 | 7 | |
| **C3** | | 38 | | 5 | 132 | 4 | | | | 44 | |
| **C4** | | 42 | 10 | 7 | 64 | | | 8 | | 141 | 2 |
| **C5** | | | 7 | | 130 | 12 | | | 4 | 31 | |
| **C6** | | 40 | | | 108 | | | 4 | | | |
| **C7** | | | 8 | | 57 | 44 | | | | 2 | |
| **C8** | | | | | 117 | | | | | | |
| **C9** | | 6 | | 4 | 47 | | | | | | |
| **C10** | | 25 | 2 | | 43 | | | 13 | 11 | 18 | |
| **C11** | | 36 | 5 | | 58 | | | | 2 | 6 | |
| **C12** | | 34 | | 10 | 45 | | | | | 7 | |
| **C13** | | 18 | | | 52 | | | 9 | | 8 | |
| **C14** | | 17 | 8 | 4 | 30 | 14 | | | 4 | 24 | |
| **C15** | | 23 | 5 | | 43 | | | | | 6 | 4 |
| **C16** | | | | | 54 | | 3 | | | | |
| **C17** | | | | | 59 | | | 7 | | | 2 |
| **C18** | | 10 | 16 | | 40 | | 3 | 2 | 14 | | |
| **C19** | | 41 | | 5 | 11 | | 5 | | | 9 | |
| **C20** | | 10 | 2 | | 47 | 11 | | | | | |
| **C21** | 18 | 10 | | 7 | 39 | | | | | | |
| **C22** | | 34 | 7 | | 25 | | | | | | |
| **C23** | | | | 7 | 15 | 4 | 13 | | | 18 | |
| **C24** | | | | | 55 | | | | | | 4 |
| **C25** | | 3 | | | 42 | | | | | 10 | |
| **C26** | | 17 | | 4 | 14 | | | | | | 6 |
| **C27** | | | | | 19 | | | | | | |
| **C28** | | 17 | 3 | 18 | 8 | | | 4 | | | |
| **C29** | | 15 | | | 33 | | | | | | |
| **C30** | | | | | 22 | | | | | 8 | |
| **C31** | 4 | 6 | | | 7 | | | | | 20 | |
| **C32** | | 17 | | | 22 | | | | 4 | 2 | |
| **C33** | | | 2 | 4 | 30 | | | | 2 | 8 | 2 |
| **C34** | | | 12 | | 51 | | | 16 | | | |
| **C35** | | 3 | | | 39 | | | | | 18 | |
| **C36** | | 6 | | | 6 | | | | | 8 | |
| **C37** | | | | | 22 | | | | | 6 | |
| **C38** | | 54 | | | | | | | | | |
| **C39** | | | 22 | | 27 | | | | | | |
| **C40** | | | | | 18 | | 3 | | | 3 | |
| **C41** | | 17 | | | 8 | | | | | | |
| **C42** | | | 13 | 5 | | 3 | | | | | |
| **C43** | | | | | | 6 | | | | | 8 |
| **C44** | | 17 | | | 4 | | | | 4 | | |
| **C45** | | | | 3 | 7 | | | | | | |
| **C46** | | | | | 23 | | | | | 18 | 2 |
| **C47** | | | | | 4 | | | | | 13 | |
| **C48** | | 2 | 7 | | 19 | | | | | 2 | |
| **C49** | | 4 | | 9 | 3 | | | | | | |
| **C50** | | | 5 | | 3 | | | | | | |
| **C51** | | | | | 5 | | | | | | |
| **C52** | | | 8 | | 8 | | | 2 | | | |
| **C53** | | | 11 | | | | | | | | |

*Participations of agents in projects with specific technological domains (adj. res.)*

| MKV2_AS | Others | Bio-Technologies | Organic Chemistry | Geotermal Sciences and Biomasses | ICT - Multimedia | Mechanics Engineering | Multi Disciplinary | Nanotechnologies | New Materials | Optoelectronic | NA |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **C0** | -1,02 | -1,75 | 1,55 | -2,23 | 0,41 | 11,68 | 2,66 | -2,15 | -0,20 | -5,07 | -1,19 |
| **C1** | -1,36 | -3,61 | -1,02 | -1,48 | -1,33 | 6,85 | 1,33 | 6,42 | -2,08 | 1,08 | -1,59 |
| **C2** | -1,04 | -5,95 | -3,00 | 1,02 | 8,86 | -3,00 | -1,36 | 1,22 | 1,07 | -3,58 | -1,22 |
| **C3** | -1,16 | 0,63 | -3,33 | -0,38 | 1,45 | -1,99 | -1,51 | -2,44 | -1,77 | 3,47 | -1,35 |
| **C4** | -1,29 | -0,10 | -0,70 | -0,09 | -10,73 | -3,71 | -1,68 | 0,51 | -1,98 | 20,45 | -0,08 |
| **C5** | -1,05 | -5,96 | -0,46 | -2,29 | 4,51 | 1,39 | -1,36 | -2,21 | 1,06 | 1,92 | -1,22 |
| **C6** | -0,95 | 3,74 | -2,73 | -2,07 | 4,18 | -2,72 | -1,23 | 0,14 | -1,45 | -4,71 | -1,11 |
| **C7** | -0,81 | -4,59 | 1,40 | -1,76 | -0,67 | 18,21 | -1,05 | -1,70 | -1,23 | -3,42 | -0,94 |
| **C8** | -0,83 | -4,71 | -2,38 | -1,81 | 10,04 | -2,37 | -1,08 | -1,75 | -1,27 | -4,12 | -0,97 |
| **C9** | -0,57 | -1,05 | -1,65 | 2,08 | 4,27 | -1,64 | -0,75 | -1,21 | -0,88 | -2,85 | -0,67 |
| **C10** | -0,81 | 2,01 | -1,40 | -1,77 | -3,47 | -2,32 | -1,05 | 6,34 | 8,05 | 1,23 | -0,95 |
| **C11** | -0,79 | 5,24 | 0,09 | -1,73 | -0,06 | -2,27 | -1,03 | -1,67 | 0,52 | -2,14 | -0,92 |
| **C12** | -0,75 | 5,44 | -2,15 | 4,81 | -1,52 | -2,14 | -0,97 | -1,58 | -1,14 | -1,52 | -0,87 |
| **C13** | -0,71 | 1,34 | -2,05 | -1,55 | 1,00 | -2,04 | -0,92 | 4,80 | -1,09 | -0,90 | -0,83 |
| **C14** | -0,77 | 0,36 | 1,69 | 0,84 | -5,07 | 4,63 | -1,00 | -1,62 | 2,38 | 3,55 | -0,90 |
| **C15** | -0,69 | 3,23 | 0,74 | -1,50 | -0,26 | -1,97 | -0,89 | -1,45 | -1,05 | -1,36 | 4,34 |
| **C16** | -0,57 | -3,26 | -1,65 | -1,25 | 6,15 | -1,64 | 3,38 | -1,21 | -0,88 | -2,85 | -0,67 |
| **C17** | -0,63 | -3,57 | -1,80 | -1,37 | 5,39 | -1,80 | -0,81 | 4,20 | -0,96 | -3,12 | 2,07 |
| **C18** | -0,70 | -0,97 | 6,45 | -1,54 | -1,39 | -2,02 | 2,48 | -0,07 | 12,44 | -3,49 | -0,82 |
| **C19** | -0,64 | 9,90 | -1,84 | 2,33 | -6,66 | -1,84 | 5,34 | -1,35 | -0,98 | 0,09 | -0,75 |
| **C20** | -0,64 | -0,29 | -0,67 | -1,39 | 2,15 | 4,60 | -0,83 | -1,34 | -0,97 | -3,16 | -0,74 |
| **C21** | 27,55 | -0,49 | -1,88 | 3,69 | -0,31 | -1,88 | -0,85 | -1,38 | -1,00 | -3,26 | -0,77 |
| **C22** | -0,62 | 8,13 | 2,42 | -1,35 | -2,73 | -1,77 | -0,80 | -1,30 | -0,94 | -3,07 | -0,72 |
| **C23** | -0,57 | -3,26 | -1,65 | 4,57 | -4,30 | 0,94 | 17,15 | -1,21 | -0,88 | 4,46 | -0,67 |
| **C24** | -0,58 | -3,32 | -1,68 | -1,28 | 6,02 | -1,67 | -0,76 | -1,23 | -0,89 | -2,90 | 5,32 |
| **C25** | -0,56 | -2,08 | -1,62 | -1,23 | 3,28 | -1,61 | -0,73 | -1,19 | -0,86 | 1,33 | -0,66 |
| **C26** | -0,49 | 4,60 | -1,40 | 2,86 | -2,63 | -1,39 | -0,63 | -1,02 | -0,74 | -2,41 | 10,21 |
| **C27** | -0,33 | -1,88 | -0,95 | -0,72 | 3,99 | -0,95 | -0,43 | -0,69 | -0,50 | -1,64 | -0,39 |
| **C28** | -0,54 | 3,62 | 0,52 | 14,80 | -5,50 | -1,54 | -0,70 | 2,54 | -0,82 | -2,67 | -0,63 |
| **C29** | -0,53 | 3,02 | -1,51 | -1,15 | 2,00 | -1,51 | -0,68 | -1,11 | -0,80 | -2,61 | -0,61 |
| **C30** | -0,41 | -2,36 | -1,19 | -0,91 | 2,08 | -1,19 | -0,54 | -0,87 | -0,63 | 2,40 | -0,48 |
| **C31** | 8,36 | 0,11 | -1,33 | -1,01 | -4,37 | -1,32 | -0,60 | -0,97 | -0,70 | 7,76 | -0,54 |
| **C32** | -0,51 | 4,14 | -1,46 | -1,11 | -0,76 | -1,46 | -0,66 | -1,07 | 4,50 | -1,62 | -0,59 |
| **C33** | -0,53 | -2,99 | -0,11 | 2,47 | 1,12 | -1,51 | -0,68 | -1,11 | 1,75 | 0,92 | 2,71 |
| **C34** | -0,68 | -3,85 | 4,64 | -1,48 | 1,82 | -1,94 | -0,88 | 10,31 | -1,03 | -3,37 | -0,79 |
| **C35** | -0,59 | -2,27 | -1,69 | -1,29 | 1,65 | -1,69 | -0,76 | -1,24 | -0,90 | 4,20 | -0,69 |
| **C36** | -0,34 | 1,79 | -0,97 | -0,74 | -2,21 | -0,97 | -0,44 | -0,71 | -0,52 | 3,78 | -0,40 |
| **C37** | -0,40 | -2,28 | -1,15 | -0,88 | 2,57 | -1,15 | -0,52 | -0,84 | -0,61 | 1,47 | -0,47 |
| **C38** | -0,56 | 17,25 | -1,60 | -1,22 | -8,10 | -1,60 | -0,72 | -1,18 | -0,85 | -2,77 | -0,65 |
| **C39** | -0,53 | -3,02 | 13,75 | -1,16 | 0,09 | -1,52 | -0,69 | -1,12 | -0,81 | -2,64 | -0,62 |
| **C40** | -0,37 | -2,11 | -1,07 | -0,81 | 2,02 | -1,06 | 5,85 | -0,78 | -0,57 | 0,03 | -0,43 |
| **C41** | -0,38 | 7,26 | -1,09 | -0,83 | -2,27 | -1,08 | -0,49 | -0,80 | -0,58 | -1,88 | -0,44 |
| **C42** | -0,35 | -1,97 | 12,74 | 6,06 | -5,03 | 2,19 | -0,45 | -0,73 | -0,53 | -1,72 | -0,41 |
| **C43** | -0,28 | -1,61 | -0,81 | -0,62 | -4,10 | 6,97 | -0,37 | -0,60 | -0,43 | -1,41 | 24,18 |
| **C44** | -0,38 | 7,26 | -1,09 | -0,83 | -3,88 | -1,08 | -0,49 | -0,80 | 6,49 | -1,88 | -0,44 |
| **C45** | -0,24 | -1,36 | -0,69 | 5,40 | 0,99 | -0,68 | -0,31 | -0,50 | -0,37 | -1,19 | -0,28 |
| **C46** | -0,50 | -2,83 | -1,43 | -1,09 | -0,13 | -1,43 | -0,65 | -1,05 | -0,76 | 5,93 | 2,93 |
| **C47** | -0,31 | -1,77 | -0,90 | -0,68 | -2,57 | -0,89 | -0,41 | -0,66 | -0,48 | 8,07 | -0,36 |
| **C48** | -0,41 | -1,35 | 5,00 | -0,91 | 0,98 | -1,19 | -0,54 | -0,87 | -0,63 | -0,95 | -0,48 |
| **C49** | -0,30 | 1,05 | -0,87 | 13,40 | -2,88 | -0,87 | -0,39 | -0,64 | -0,46 | -1,50 | -0,35 |
| **C50** | -0,21 | -1,22 | 7,93 | -0,47 | -0,97 | -0,61 | -0,28 | -0,45 | -0,33 | -1,06 | -0,25 |
| **C51** | -0,17 | -0,96 | -0,49 | -0,37 | 2,05 | -0,48 | -0,22 | -0,36 | -0,26 | -0,84 | -0,20 |
| **C52** | -0,32 | -1,83 | 8,21 | -0,70 | -0,86 | -0,92 | -0,42 | 2,37 | -0,49 | -1,59 | -0,38 |
| **C53** | -0,25 | -1,43 | 15,32 | -0,55 | -3,63 | -0,72 | -0,33 | -0,53 | -0,38 | -1,25 | -0,29 |

**Mkv2_AS**




*Graph of the partition. Node are proportional to the number of agents that belong to the corresponding community. Edges represent overlaps between communities and they are proportional to the number of nodes involved. Fruchterman Reingold layout.*

## a.6. MKV2_DS

*Typology of agents (obs. freq)*

| MKV2_DS | GI | TA | CC | SC | LGI | E | R | BS | KIBS |
|---|---|---|---|---|---|---|---|---|---|
| C0 | | 1 | 2 | 1 | 2 | 14 | 4 | 1 | 15 |
| C1 | 2 | 2 | 1 | | 11 | 11 | 2 | | 16 |
| C2 | 4 | 2 | 3 | | 6 | 3 | 6 | 1 | 22 |
| C3 | | | 1 | | 5 | 6 | 5 | | 27 |
| C4 | 2 | 1 | | | 6 | 17 | 6 | | 18 |
| C5 | 1 | | 1 | | 5 | 8 | 2 | | 14 |
| C6 | | | 2 | | 6 | 9 | 6 | 1 | 8 |
| C7 | 1 | 3 | 2 | | 5 | 6 | 3 | | 16 |
| C8 | | | 2 | | 2 | 12 | | | 4 |
| C9 | | | | 1 | | 2 | 1 | 1 | 20 |
| C10 | 3 | | | | 5 | 6 | 1 | 1 | 1 |
| C11 | 1 | | 4 | | 4 | | 2 | | 11 |
| C12 | | 1 | 1 | | 3 | 3 | 1 | | 12 |
| C13 | | | 1 | | 5 | 3 | 1 | | 9 |
| C14 | | | 12 | | | 1 | | | 4 |
| C15 | | | 1 | | 1 | 2 | 5 | | 7 |
| C16 | | | | 1 | | 3 | 1 | | 8 |
| C17 | 1 | | | | 1 | 6 | 2 | | 8 |
| C18 | | | | | 2 | 4 | 1 | 1 | 9 |
| C19 | | | 1 | | 3 | 1 | 2 | | 6 |
| C20 | 1 | | 2 | | 1 | 1 | 3 | 2 | 3 |
| C21 | | | | | | 3 | 4 | | 4 |
| C22 | | | | | 1 | 4 | 3 | | 4 |
| C23 | | | | | 4 | 6 | 2 | | 5 |
| C24 | 1 | 2 | | | | 4 | 2 | | 3 |
| C25 | | 1 | 1 | | | | 1 | | 3 |
| C26 | 1 | | | | | | | | 3 |
| C27 | | | | | 2 | 2 | 2 | | 4 |
| C28 | | | | | | 6 | | | 5 |
| C29 | | | | | 2 | | 1 | 1 | 4 |
| C30 | 1 | | | | 1 | 3 | 1 | 1 | 3 |
| C31 | | | | | | | | | 5 |
| C32 | | | | 1 | | | 2 | | 6 |
| C33 | | | | | 2 | 2 | 1 | | 1 |
| C34 | | | 1 | | 1 | 1 | 1 | | 5 |
| C35 | | | | | | 1 | | | 5 |
| C36 | | | | | | 2 | 1 | | 3 |
| C37 | | | 1 | | | 4 | 2 | | 5 |
| C38 | | | | | | 1 | 3 | | 2 |
| C39 | 2 | | | | | 2 | 2 | | |
| C40 | 1 | | | | 1 | 3 | | | 2 |
| C41 | | | | | | | | | 3 |
| C42 | | | | | 2 | | | | 1 |
| C43 | | | | | | 1 | | | 2 |
| C44 | | | | | | 3 | | 1 | 1 |
| C45 | | | | | 3 | | 1 | | 1 |
| C46 | | | | | 1 | 1 | 1 | | 4 |
| C47 | | | | | | 1 | | | 3 |
| C48 | | 1 | 1 | | | 1 | | | 2 |
| C49 | | | | | | 1 | 1 | 1 | 1 |
| C50 | 1 | | | | 1 | 2 | | | |
| C51 | | | | | | 2 | | | 2 |
| C52 | | | | | | | 2 | | |
| C53 | | 1 | | | | | | | 2 |
| C54 | | | | | | 1 | 1 | | 4 |
| C55 | | | | | | | | | 2 |
| C56 | | | | | | 2 | 1 | | 1 |
| C57 | | | | | | 2 | 2 | | 1 |
| C58 | | | | | | | | | 3 |
| C59 | | | | | | | | | 2 |
| C60 | | | | | | 1 | | | 1 |
| C61 | | | | | | | 1 | | 1 |

*Typology of agents (adj. res)*

| MKV2_DS | GI | TA | CC | SC | LGI | E | R | BS | KIBS |
|---|---|---|---|---|---|---|---|---|---|
| C0 | -1,11 | 0,31 | 0,01 | 1,85 | -1,34 | 1,98 | -0,29 | 0,54 | -0,72 |
| C1 | 0,66 | 1,32 | -0,87 | -0,49 | 2,75 | 0,36 | -1,51 | -0,85 | -1,04 |
| C2 | 2,40 | 1,25 | 0,46 | -0,50 | 0,25 | -2,70 | 0,30 | 0,37 | 0,54 |
| C3 | -1,17 | -0,94 | -0,84 | -0,48 | -0,06 | -1,42 | -0,01 | -0,84 | 2,53 |
| C4 | 0,51 | 0,08 | -1,67 | -0,52 | 0,08 | 2,05 | 0,14 | -0,90 | -1,03 |
| C5 | 0,13 | -0,78 | -0,45 | -0,40 | 0,79 | 0,48 | -0,88 | -0,70 | 0,25 |
| C6 | -0,99 | -0,79 | 0,34 | -0,41 | 1,28 | 0,81 | 1,34 | 0,78 | -2,10 |
| C7 | -0,03 | 2,94 | 0,17 | -0,43 | 0,43 | -0,83 | -0,59 | -0,75 | 0,18 |
| C8 | -0,78 | -0,62 | 1,05 | -0,32 | -0,23 | 4,10 | -1,62 | -0,56 | -2,10 |
| C9 | -0,87 | -0,70 | -1,16 | 2,54 | -1,84 | -1,75 | -1,18 | 1,06 | 3,80 |
| C10 | 3,71 | -0,57 | -0,95 | -0,29 | 2,31 | 1,30 | -0,72 | 1,51 | -3,12 |
| C11 | 0,49 | -0,65 | 2,90 | -0,34 | 0,97 | -2,55 | -0,35 | -0,58 | 0,67 |
| C12 | -0,80 | 1,00 | -0,04 | -0,33 | 0,38 | -0,89 | -0,97 | -0,57 | 1,33 |
| C13 | -0,76 | -0,61 | 0,06 | -0,31 | 2,02 | -0,69 | -0,85 | -0,54 | 0,39 |
| C14 | -0,71 | -0,57 | 12,60 | -0,29 | -1,51 | -1,64 | -1,50 | -0,51 | -1,64 |
| C15 | -0,69 | -0,56 | 0,24 | -0,29 | -0,68 | -0,95 | 2,52 | -0,50 | 0,06 |
| C16 | -0,62 | -0,50 | -0,83 | 3,73 | -1,32 | 0,07 | -0,42 | -0,45 | 1,36 |
| C17 | 0,70 | -0,59 | -0,98 | -0,30 | -0,82 | 1,14 | -0,04 | -0,53 | 0,13 |
| C18 | -0,71 | -0,57 | -0,95 | -0,29 | 0,02 | 0,12 | -0,72 | 1,51 | 0,84 |
| C19 | -0,62 | -0,50 | 0,46 | -0,26 | 1,30 | -1,28 | 0,46 | -0,45 | 0,23 |
| C20 | 1,06 | -0,50 | 1,75 | -0,26 | -0,45 | -1,28 | 1,34 | 4,17 | -1,46 |
| C21 | -0,57 | -0,46 | -0,76 | -0,24 | -1,21 | 0,40 | 2,62 | -0,41 | -0,45 |
| C22 | -0,60 | -0,48 | -0,80 | -0,25 | -0,36 | 0,93 | 1,49 | -0,43 | -0,68 |
| C23 | -0,71 | -0,57 | -0,95 | -0,29 | 1,54 | 1,30 | 0,05 | -0,51 | -1,14 |
| C24 | 1,15 | 3,83 | -0,80 | -0,25 | -1,27 | 0,93 | 0,58 | -0,43 | -1,27 |
| C25 | -0,42 | 2,70 | 1,33 | -0,17 | -0,89 | -1,32 | 0,41 | -0,30 | 0,35 |
| C26 | 2,67 | -0,28 | -0,46 | -0,14 | -0,73 | -1,07 | -0,72 | -0,25 | 1,30 |
| C27 | -0,55 | -0,44 | -0,73 | -0,23 | 0,83 | -0,18 | 0,86 | -0,39 | -0,19 |
| C28 | -0,57 | -0,46 | -0,76 | -0,24 | -1,21 | 2,59 | -1,20 | -0,41 | 0,17 |
| C29 | -0,49 | -0,39 | -0,65 | -0,20 | 1,18 | -1,52 | 0,10 | 2,59 | 0,40 |
| C30 | 1,37 | -0,44 | -0,73 | -0,23 | -0,16 | 0,59 | -0,14 | 2,24 | -0,84 |
| C31 | -0,38 | -0,31 | -0,51 | -0,16 | -0,81 | -1,20 | -0,81 | -0,28 | 2,58 |
| C32 | -0,52 | -0,42 | -0,69 | 4,56 | -1,10 | -1,62 | 1,03 | -0,37 | 1,44 |
| C33 | -0,42 | -0,34 | -0,56 | -0,17 | 1,66 | 0,65 | 0,41 | -0,30 | -1,31 |
| C34 | -0,52 | -0,42 | 0,86 | -0,21 | -0,05 | -0,81 | -0,03 | -0,37 | 0,77 |
| C35 | -0,42 | -0,34 | -0,56 | -0,17 | -0,89 | -0,33 | -0,88 | -0,30 | 2,00 |
| C36 | -0,42 | -0,34 | -0,56 | -0,17 | -0,89 | 0,65 | 0,41 | -0,30 | 0,35 |
| C37 | -0,60 | -0,48 | 0,54 | -0,25 | -1,27 | 0,93 | 0,58 | -0,43 | -0,09 |
| C38 | -0,42 | -0,34 | -0,56 | -0,17 | -0,89 | -0,33 | 2,99 | -0,30 | -0,48 |
| C39 | 4,50 | -0,34 | -0,56 | -0,17 | -0,89 | 0,65 | 1,70 | -0,30 | -2,14 |
| C40 | 1,83 | -0,37 | -0,61 | -0,19 | 0,22 | 1,31 | -0,95 | -0,33 | -0,77 |
| C41 | -0,30 | -0,24 | -0,40 | -0,12 | -0,63 | -0,93 | -0,62 | -0,21 | 2,00 |
| C42 | -0,30 | -0,24 | -0,40 | -0,12 | 2,98 | -0,93 | -0,62 | -0,21 | -0,34 |
| C43 | -0,30 | -0,24 | -0,40 | -0,12 | -0,63 | 0,46 | -0,62 | -0,21 | 0,83 |
| C44 | -0,38 | -0,31 | -0,51 | -0,16 | -0,81 | 2,03 | -0,81 | 3,43 | -1,04 |
| C45 | -0,38 | -0,31 | -0,51 | -0,16 | 3,38 | -1,20 | 0,61 | -0,28 | -1,04 |
| C46 | -0,46 | -0,37 | -0,61 | -0,19 | 0,22 | -0,51 | 0,24 | -0,33 | 0,76 |
| C47 | -0,34 | -0,28 | -0,46 | -0,14 | -0,73 | 0,13 | -0,72 | -0,25 | 1,30 |
| C48 | -0,38 | 3,01 | 1,56 | -0,16 | -0,81 | -0,12 | -0,81 | -0,28 | -0,14 |
| C49 | -0,34 | -0,28 | -0,46 | -0,14 | -0,73 | 0,13 | 0,86 | 3,90 | -0,73 |
| C50 | 2,67 | -0,28 | -0,46 | -0,14 | 0,83 | 1,33 | -0,72 | -0,25 | -1,74 |
| C51 | -0,34 | -0,28 | -0,46 | -0,14 | -0,73 | 1,33 | -0,72 | -0,25 | 0,28 |
| C52 | -0,24 | -0,20 | -0,32 | -0,10 | -0,51 | -0,76 | 3,95 | -0,17 | -1,23 |
| C53 | -0,30 | 4,04 | -0,40 | -0,12 | -0,63 | -0,93 | -0,62 | -0,21 | 0,83 |
| C54 | -0,42 | -0,34 | -0,56 | -0,17 | -0,89 | -0,33 | 0,41 | -0,30 | 1,18 |
| C55 | -0,24 | -0,20 | -0,32 | -0,10 | -0,51 | -0,76 | -0,51 | -0,17 | 1,63 |
| C56 | -0,34 | -0,28 | -0,46 | -0,14 | -0,73 | 1,33 | 0,86 | -0,25 | -0,73 |
| C57 | -0,38 | -0,31 | -0,51 | -0,16 | -0,81 | 0,95 | 2,02 | -0,28 | -1,04 |
| C58 | -0,30 | -0,24 | -0,40 | -0,12 | -0,63 | -0,93 | -0,62 | -0,21 | 2,00 |
| C59 | -0,24 | -0,20 | -0,32 | -0,10 | -0,51 | -0,76 | -0,51 | -0,17 | 1,63 |
| C60 | -0,24 | -0,20 | -0,32 | -0,10 | -0,51 | 0,94 | -0,51 | -0,17 | 0,20 |
| C61 | -0,24 | -0,20 | -0,32 | -0,10 | -0,51 | -0,76 | 1,72 | -0,17 | 0,20 |

*Participations of agents in waves (obs. freq)*

| MKV2_DS | 2002_171 | 2002_172 | 2002_ITT | 2004_171 | 2004_171A | 2005_171 | 2006_VIN | 2007_171 | 2008_171 |
|---|---|---|---|---|---|---|---|---|---|
| C0 | 39 | 12 | 46 | 27 | 2 | 36 | 3 | 10 | |
| C1 | 60 | 8 | 31 | | 7 | 114 | | 3 | |
| C2 | 20 | 17 | 37 | 8 | 18 | 135 | | 2 | |
| C3 | | 27 | 107 | 5 | | 146 | | 6 | |
| C4 | 41 | 14 | 187 | 2 | 17 | 42 | 4 | 6 | |
| C5 | 22 | 9 | 96 | | | 115 | | | |
| C6 | 11 | 20 | 62 | | 26 | 91 | | | |
| C7 | 21 | 15 | 63 | | 4 | 71 | | 3 | |
| C8 | 24 | | 2 | 33 | | 19 | | | |
| C9 | 163 | | | | | | | 3 | |
| C10 | 3 | 6 | 6 | 4 | | 38 | | | |
| C11 | 18 | | 20 | 4 | 9 | 59 | | | |
| C12 | 8 | 8 | 30 | | 10 | 49 | | | |
| C13 | | 5 | 68 | | | 48 | | | |
| C14 | | | 42 | 8 | 4 | 52 | | | |
| C15 | 8 | 5 | 60 | | | 29 | | 4 | |
| C16 | 71 | | 21 | | | 10 | | | 2 |
| C17 | | | 84 | | | | | 2 | |
| C18 | 6 | 8 | 4 | | | 44 | 2 | | |
| C19 | | 5 | 35 | | 9 | 28 | | | |
| C20 | | 10 | 20 | 8 | 8 | 30 | | | |
| C21 | | | 39 | 5 | | 14 | | | |
| C22 | 7 | 20 | 9 | | | 22 | | 2 | |
| C23 | 43 | 18 | 12 | | 14 | 7 | | 6 | |
| C24 | 19 | | 38 | | | | | | |
| C25 | 5 | | | | 4 | 5 | | 4 | |
| C26 | | 8 | | | | 3 | | 8 | |
| C27 | 16 | | 18 | | | 16 | | | |
| C28 | 7 | | 12 | 6 | | 7 | 2 | | |
| C29 | | 8 | | | 7 | 19 | | | |
| C30 | | | 7 | 16 | 2 | | 3 | 5 | |
| C31 | | 4 | 7 | | | 5 | | 2 | |
| C32 | 4 | | 38 | | | | | | |
| C33 | | 2 | 6 | | | 9 | | 2 | |
| C34 | | | 17 | | | 32 | | | |
| C35 | | 3 | | | | 11 | 7 | | |
| C36 | | 3 | 8 | | | 3 | 6 | | |
| C37 | 16 | | 32 | 12 | | | | | |
| C38 | | | 30 | | | | | | |
| C39 | | 5 | 2 | 2 | | 6 | | | 2 |
| C40 | | 22 | 6 | | | | | | |
| C41 | 4 | | | | | | | 4 | |
| C42 | | | 6 | | | 2 | | 2 | |
| C43 | | | | | | 4 | | 4 | |
| C44 | | | | | 5 | 3 | 5 | | |
| C45 | | 2 | 9 | | | 4 | | | |
| C46 | | 5 | 3 | | | 12 | | 4 | |
| C47 | | 8 | 6 | 4 | | 2 | | | |
| C48 | | | | | | 6 | | 8 | |
| C49 | | 3 | | 3 | | | | 4 | |
| C50 | | | 8 | | | 5 | 2 | | |
| C51 | | | 2 | 8 | | | 3 | 3 | |
| C52 | | | | | | | 2 | 2 | |
| C53 | | | | | | 4 | | 4 | |
| C54 | 4 | 17 | | | | | | 2 | |
| C55 | | 3 | | | | | | 2 | |
| C56 | 4 | | 9 | | | | | | |
| C57 | | 13 | | | | | | | |
| C58 | | 3 | 4 | | | | | | |
| C59 | | 3 | | | 2 | | | | |
| C60 | | 3 | | | 2 | | | | |
| C61 | | 3 | 4 | | | | | | |

*Participations of agents in waves (adj. res.)*

| MKV2_DS | 2002_171 | 2002_172 | 2002_ITT | 2004_171 | 2004_171A | 2005_171 | 2006_VIN | 2007_171 | 2008_171 |
|---|---|---|---|---|---|---|---|---|---|
| C0 | 2,50 | -0,51 | -1,86 | 8,31 | -1,80 | -3,53 | 1,08 | 2,66 | -0,42 |
| C1 | 4,80 | -2,44 | -6,16 | -3,03 | -0,40 | 5,98 | -1,50 | -1,20 | -0,48 |
| C2 | -3,12 | -0,41 | -5,78 | -0,31 | 3,36 | 8,15 | -1,55 | -1,74 | -0,49 |
| C3 | -7,60 | 0,93 | 1,52 | -1,89 | -3,43 | 6,54 | -1,73 | -0,59 | -0,55 |
| C4 | -1,26 | -2,32 | 10,60 | -3,01 | 1,77 | -7,61 | 0,64 | -0,78 | -0,57 |
| C5 | -2,87 | -2,47 | 2,37 | -3,16 | -3,11 | 5,02 | -1,57 | -2,61 | -0,50 |
| C6 | -4,24 | 0,92 | -1,02 | -2,94 | 6,96 | 3,33 | -1,45 | -2,43 | -0,46 |
| C7 | -1,39 | 0,31 | 0,83 | -2,68 | -1,00 | 2,11 | -1,33 | -0,77 | -0,42 |
| C8 | 3,74 | -2,60 | -5,73 | 18,10 | -1,73 | -1,61 | -0,87 | -1,45 | -0,28 |
| C9 | 29,96 | -3,84 | -9,17 | -2,60 | -2,55 | -9,19 | -1,28 | -0,65 | -0,41 |
| C10 | -2,16 | 0,75 | -3,60 | 1,31 | -1,48 | 5,48 | -0,74 | -1,24 | -0,24 |
| C11 | 0,23 | -3,11 | -3,30 | -0,06 | 2,59 | 4,71 | -1,04 | -1,73 | -0,33 |
| C12 | -2,28 | -0,09 | -0,92 | -2,05 | 3,27 | 3,06 | -1,01 | -1,69 | -0,32 |
| C13 | -4,80 | -1,55 | 5,58 | -2,20 | -2,17 | 1,63 | -1,09 | -1,82 | -0,35 |
| C14 | -4,48 | -3,05 | 1,53 | 2,09 | 0,08 | 3,61 | -1,02 | -1,70 | -0,33 |
| C15 | -2,31 | -1,22 | 5,31 | -2,06 | -2,02 | -1,21 | -1,02 | 0,78 | -0,33 |
| C16 | 15,01 | -3,02 | -2,76 | -2,04 | -2,00 | -5,11 | -1,01 | -1,68 | 6,07 |
| C17 | -4,03 | -2,74 | 12,97 | -1,85 | -1,82 | -6,55 | -0,92 | -0,16 | -0,29 |
| C18 | -1,38 | 1,39 | -4,55 | -1,59 | -1,56 | 6,17 | 1,82 | -1,31 | -0,25 |
| C19 | -3,81 | -0,45 | 2,40 | -1,75 | 3,82 | 0,67 | -0,86 | -1,44 | -0,28 |
| C20 | -3,78 | 1,73 | -1,20 | 3,14 | 3,25 | 1,25 | -0,86 | -1,43 | -0,27 |
| C21 | -3,30 | -2,24 | 5,64 | 1,97 | -1,49 | -1,42 | -0,75 | -1,25 | -0,24 |
| C22 | -0,84 | 7,39 | -2,95 | -1,54 | -1,51 | 0,64 | -0,76 | 0,37 | -0,24 |
| C23 | 7,66 | 3,81 | -4,47 | -2,00 | 5,62 | -5,57 | -0,99 | 2,18 | -0,32 |
| C24 | 3,72 | -2,22 | 5,50 | -1,50 | -1,48 | -5,31 | -0,74 | -1,24 | -0,24 |
| C25 | 1,43 | -1,24 | -2,97 | -0,84 | 4,23 | -0,46 | -0,42 | 5,26 | -0,13 |
| C26 | -1,88 | 5,56 | -3,05 | -0,86 | -0,85 | -1,59 | -0,43 | 10,87 | -0,14 |
| C27 | 3,22 | -2,08 | 0,50 | -1,40 | -1,38 | -0,13 | -0,69 | -1,16 | -0,22 |
| C28 | 0,81 | -1,71 | 0,32 | 4,28 | -1,14 | -1,53 | 2,99 | -0,95 | -0,18 |
| C29 | -2,52 | 3,41 | -4,08 | -1,16 | 5,31 | 2,88 | -0,57 | -0,95 | -0,18 |
| C30 | -2,48 | -1,69 | -1,42 | 13,58 | 0,75 | -4,03 | 4,86 | 4,56 | -0,18 |
| C31 | -1,83 | 2,27 | 0,56 | -0,84 | -0,83 | -0,46 | -0,42 | 2,28 | -0,13 |
| C32 | -1,09 | -1,90 | 8,02 | -1,29 | -1,26 | -4,55 | -0,64 | -1,06 | -0,20 |
| C33 | -1,88 | 0,43 | -0,11 | -0,86 | -0,85 | 1,35 | -0,43 | 2,18 | -0,14 |
| C34 | -3,03 | -2,06 | 0,30 | -1,39 | -1,37 | 4,87 | -0,69 | -1,15 | -0,22 |
| C35 | -1,97 | 1,10 | -3,21 | -0,91 | -0,89 | 1,91 | 15,39 | -0,75 | -0,14 |
| C36 | -1,93 | 1,19 | 0,70 | -0,89 | -0,87 | -1,70 | 13,48 | -0,73 | -0,14 |
| C37 | 2,39 | -2,28 | 3,43 | 6,68 | -1,51 | -5,45 | -0,76 | -1,27 | -0,24 |
| C38 | -2,36 | -1,61 | 7,88 | -1,09 | -1,07 | -3,84 | -0,54 | -0,90 | -0,17 |
| C39 | -1,78 | 3,31 | -1,85 | 1,74 | -0,80 | 0,22 | -0,40 | -0,67 | 15,51 |
| C40 | -2,28 | 13,95 | -1,28 | -1,05 | -1,03 | -3,71 | -0,52 | -0,87 | -0,17 |
| C41 | 2,69 | -0,83 | -1,98 | -0,56 | -0,55 | -1,98 | -0,28 | 8,45 | -0,09 |
| C42 | -1,36 | -0,93 | 1,84 | -0,63 | -0,61 | -0,87 | -0,31 | 3,47 | -0,10 |
| C43 | -1,22 | -0,83 | -1,98 | -0,56 | -0,55 | 1,04 | -0,28 | 8,45 | -0,09 |
| C44 | -1,55 | -1,06 | -2,52 | -0,71 | 6,73 | -0,75 | 14,02 | -0,59 | -0,11 |
| C45 | -1,67 | 0,79 | 2,26 | -0,77 | -0,75 | -0,51 | -0,38 | -0,63 | -0,12 |
| C46 | -2,11 | 2,37 | -2,12 | -0,97 | -0,95 | 1,80 | -0,48 | 4,36 | -0,15 |
| C47 | -1,93 | 5,35 | -0,26 | 3,84 | -0,87 | -2,18 | -0,44 | -0,73 | -0,14 |
| C48 | -1,61 | -1,10 | -2,61 | -0,74 | -0,73 | 0,80 | -0,37 | 12,88 | -0,12 |
| C49 | -1,36 | 2,60 | -2,21 | 4,38 | -0,61 | -2,21 | -0,31 | 7,46 | -0,10 |
| C50 | -1,67 | -1,13 | 1,70 | -0,77 | -0,75 | 0,04 | 4,97 | -0,63 | -0,12 |
| C51 | -1,72 | -1,17 | -1,73 | 9,76 | -0,78 | -2,80 | 7,38 | 4,08 | -0,13 |
| C52 | -0,86 | -0,58 | -1,40 | -0,40 | -0,39 | -1,40 | 10,15 | 5,98 | -0,06 |
| C53 | -1,22 | -0,83 | -1,98 | -0,56 | -0,55 | 1,04 | -0,28 | 8,45 | -0,09 |
| C54 | 0,24 | 11,80 | -3,35 | -0,95 | -0,93 | -3,36 | -0,47 | 1,85 | -0,15 |
| C55 | -0,96 | 4,33 | -1,56 | -0,44 | -0,43 | -1,56 | -0,22 | 5,27 | -0,07 |
| C56 | 1,51 | -1,06 | 2,81 | -0,71 | -0,70 | -2,52 | -0,35 | -0,59 | -0,11 |
| C57 | -1,55 | 12,36 | -2,52 | -0,71 | -0,70 | -2,52 | -0,35 | -0,59 | -0,11 |
| C58 | -1,14 | 3,44 | 1,38 | -0,52 | -0,51 | -1,85 | -0,26 | -0,43 | -0,08 |
| C59 | -0,96 | 4,33 | -1,56 | -0,44 | 4,35 | -1,56 | -0,22 | -0,37 | -0,07 |
| C60 | -0,96 | 4,33 | -1,56 | -0,44 | 4,35 | -1,56 | -0,22 | -0,37 | -0,07 |
| C61 | -1,14 | 3,44 | 1,38 | -0,52 | -0,51 | -1,85 | -0,26 | -0,43 | -0,08 |

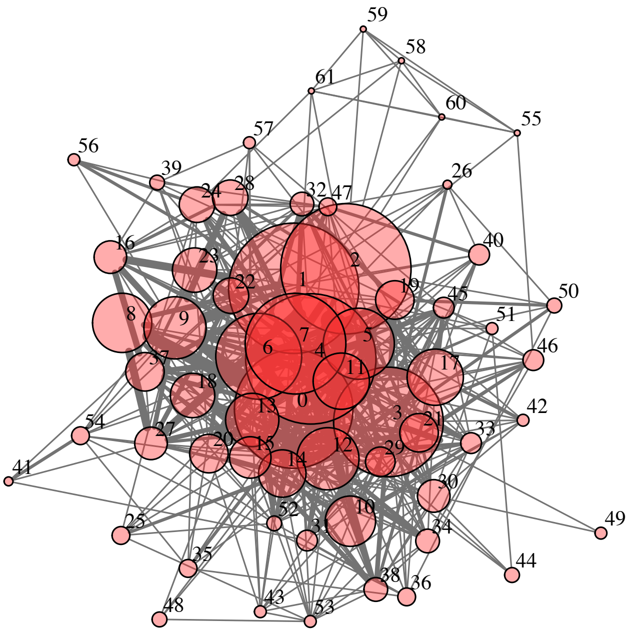
*Participations of agents in projects with specific technological domains (obs. freq)*

| MKV2_DS | Others | Bio-Technologies | Organic Chemistry | Geotermal Sciences and Biomasses | ICT - Multimedia | Mechanics Engineering | Multi Disciplinary | Nanotechnologies | New Materials | Optoelectronic | NA |
|---|---|---|---|---|---|---|---|---|---|---|---|
| C0 | | 19 | 12 | | 98 | 39 | 5 | | 2 | | |
| C1 | | | 8 | 7 | 178 | 12 | | 7 | 4 | 7 | |
| C2 | | | 13 | | 136 | 26 | 10 | 17 | | 35 | |
| C3 | | 38 | | 5 | 171 | 8 | | | | 69 | |
| C4 | | 39 | 10 | 7 | 91 | | | 8 | | 156 | 2 |
| C5 | | | | | 150 | 12 | | | 4 | 76 | |
| C6 | | 35 | 2 | | 103 | 22 | | 15 | 13 | 20 | |
| C7 | | 57 | | | 116 | | | 4 | | | |
| C8 | | | 8 | | 28 | 40 | | | | 2 | |
| C9 | | | 3 | | 163 | | | | | | |
| C10 | | 6 | | 4 | 47 | | | | | | |
| C11 | | | | 4 | 61 | 16 | | 16 | | 13 | |
| C12 | | 23 | | 10 | 65 | | | | | 7 | |
| C13 | | 68 | 5 | | 40 | | | | 2 | 6 | |
| C14 | | 34 | 8 | 4 | 22 | 14 | | 2 | 4 | 18 | |
| C15 | | 48 | 5 | | 43 | | | | | 6 | 4 |
| C16 | | | | | 95 | | | 7 | | | 2 |
| C17 | | 20 | | | | | | | | 66 | |
| C18 | | | | | 61 | | 3 | | | | |
| C19 | | 15 | | | 45 | | | 9 | | 8 | |
| C20 | | 10 | 16 | | 31 | | 3 | 2 | 14 | | |
| C21 | | 41 | | 5 | 4 | | 5 | | | 3 | |
| C22 | | 10 | 2 | | 37 | 11 | | | | | |
| C23 | 18 | 10 | | 14 | 58 | | | | | | |
| C24 | | 34 | 2 | | 21 | | | | | | |
| C25 | | | | 4 | 10 | | | | | | 4 |
| C26 | | | | | 19 | | | | | | |
| C27 | | | | | 34 | 16 | | | | | |
| C28 | | | | | 28 | | | | | | 6 |
| C29 | | | | 7 | 19 | | 8 | | | | |
| C30 | | | 3 | 18 | 8 | | | 4 | | | |
| C31 | | 3 | | | 8 | | | | | 7 | |
| C32 | | 15 | | | 27 | | | | | | |
| C33 | 4 | 6 | | | 7 | | | | | 2 | |
| C34 | | 17 | | | 26 | | | | 4 | 2 | |
| C35 | | 3 | | | 18 | | | | | | |
| C36 | | 6 | | | 6 | | | | | 8 | |
| C37 | | | 12 | | 32 | | | 16 | | | |
| C38 | | 30 | | | | | | | | | |
| C39 | | | 7 | | 6 | | | | | 2 | 2 |
| C40 | | | | | 22 | | | | | 6 | |
| C41 | | | | | 8 | | | | | | |
| C42 | | | | | 4 | | | | | 6 | |
| C43 | | | | | 4 | | | | 4 | | |
| C44 | | | 5 | 5 | | 3 | | | | | |
| C45 | | | 2 | | | | | | | 13 | |
| C46 | | | | | 18 | | 3 | | | 3 | |
| C47 | | | | 4 | 8 | | | | 2 | 6 | |
| C48 | | | | | | 6 | | | | | 8 |
| C49 | | | | 3 | 7 | | | | | | |
| C50 | | | | | 7 | | | | | 8 | |
| C51 | | | 14 | | 2 | | | | | | |
| C52 | | | | | 2 | | | | | | 2 |
| C53 | | | | | 4 | | | | | | 4 |
| C54 | | 17 | | | 6 | | | | | | |
| C55 | | | | | 5 | | | | | | |
| C56 | | 2 | 7 | | 4 | | | | | | |
| C57 | | | 13 | | | | | | | | |
| C58 | | | | | 7 | | | | | | |
| C59 | | | | 2 | 3 | | | | | | |
| C60 | | | 2 | | 3 | | | | | | |
| C61 | | 4 | | | 3 | | | | | | |

*Participations of agents in projects with specific technological domains (adj. res.)*

| MKV2_DS | Others | Bio-Technologies | Organic Chemistry | Geotermal Sciences and Biomasses | ICT - Multimedia | Mechanics Engineering | Multi Disciplinary | Nanotechnologies | New Materials | Optoelectronic | NA |
|---|---|---|---|---|---|---|---|---|---|---|---|
| C0 | -0,99 | -1,49 | 2,12 | -2,16 | 0,56 | 10,04 | 2,82 | -2,20 | -0,17 | -5,32 | -1,23 |
| C1 | -1,12 | -6,39 | -0,21 | 0,64 | 7,98 | -0,04 | -1,46 | 0,53 | 0,70 | -4,63 | -1,40 |
| C2 | -1,16 | -6,60 | 1,35 | -2,54 | 1,10 | 3,86 | 5,60 | 4,58 | -1,81 | 0,62 | -1,44 |
| C3 | -1,29 | -0,85 | -3,54 | -0,88 | 1,72 | -2,10 | -1,68 | -2,88 | -2,02 | 5,34 | -1,61 |
| C4 | -1,35 | -1,19 | -0,62 | -0,30 | -9,17 | -4,42 | -1,75 | -0,04 | -2,10 | 19,66 | -0,37 |
| C5 | -1,17 | -6,67 | -3,21 | -2,56 | 2,59 | -0,34 | -1,52 | -2,61 | 0,53 | 8,46 | -1,46 |
| C6 | -1,09 | 0,80 | -2,24 | -2,38 | -1,45 | 3,30 | -1,41 | 4,27 | 6,49 | -1,70 | -1,35 |
| C7 | -1,00 | 6,69 | -2,72 | -2,17 | 3,17 | -3,26 | -1,29 | -0,28 | -1,55 | -5,36 | -1,24 |
| C8 | -0,65 | -3,71 | 2,97 | -1,43 | -3,22 | 18,02 | -0,85 | -1,45 | -1,02 | -2,84 | -0,81 |
| C9 | -0,96 | -5,47 | -1,39 | -2,10 | 11,68 | -3,16 | -1,25 | -2,14 | -1,50 | -5,18 | -1,20 |
| C10 | -0,56 | -0,91 | -1,52 | 2,21 | 4,35 | -1,82 | -0,72 | -1,24 | -0,87 | -2,99 | -0,69 |
| C11 | -0,78 | -4,42 | -2,13 | 0,78 | 0,33 | 4,27 | -1,01 | 8,01 | -1,21 | -0,50 | -0,97 |
| C12 | -0,76 | 2,09 | -2,08 | 4,68 | 1,66 | -2,49 | -0,99 | -1,69 | -1,18 | -2,06 | -0,95 |
| C13 | -0,82 | 13,05 | 0,17 | -1,79 | -4,67 | -2,68 | -1,06 | -1,82 | 0,37 | -2,77 | -1,02 |
| C14 | -0,76 | 5,09 | 2,01 | 0,86 | -6,94 | 3,57 | -0,99 | -0,46 | 2,31 | 1,09 | -0,95 |
| C15 | -0,76 | 8,98 | 0,47 | -1,67 | -2,79 | -2,50 | -0,99 | -1,70 | -1,19 | -2,38 | 3,41 |
| C16 | -0,76 | -4,30 | -2,07 | -1,65 | 7,76 | -2,48 | -0,98 | 2,70 | -1,18 | -4,07 | 1,26 |
| C17 | -0,69 | 2,25 | -1,87 | -1,50 | -10,14 | -2,25 | -0,89 | -1,53 | -1,07 | 17,41 | -0,85 |
| C18 | -0,59 | -3,35 | -1,61 | -1,29 | 6,70 | -1,93 | 3,25 | -1,31 | -0,92 | -3,18 | -0,73 |
| C19 | -0,65 | 1,18 | -1,77 | -1,42 | 0,80 | -2,13 | -0,84 | 5,08 | -1,01 | -0,79 | -0,81 |
| C20 | -0,64 | -0,40 | 7,87 | -1,41 | -2,32 | -2,11 | 2,85 | 0,02 | 13,41 | -3,47 | -0,80 |
| C21 | -0,56 | 12,10 | -1,53 | 3,02 | -7,24 | -1,84 | 6,29 | -1,25 | -0,87 | -1,86 | -0,70 |
| C22 | -0,57 | 0,42 | -0,21 | -1,25 | 1,21 | 4,43 | -0,74 | -1,27 | -0,89 | -3,07 | -0,71 |
| C23 | 24,31 | -1,36 | -2,03 | 7,47 | 0,83 | -2,43 | -0,96 | -1,65 | -1,15 | -3,99 | -0,92 |
| C24 | -0,56 | 9,62 | -0,13 | -1,22 | -2,60 | -1,82 | -0,72 | -1,24 | -0,87 | -2,99 | -0,69 |
| C25 | -0,31 | -1,77 | -0,85 | 5,38 | 0,14 | -1,02 | -0,40 | -0,69 | -0,49 | -1,67 | 10,08 |
| C26 | -0,32 | -1,82 | -0,87 | -0,70 | 4,04 | -1,05 | -0,42 | -0,71 | -0,50 | -1,72 | -0,40 |
| C27 | -0,52 | -2,96 | -1,42 | -1,14 | 2,01 | 8,33 | -0,68 | -1,16 | -0,81 | -2,80 | -0,65 |
| C28 | -0,43 | -2,44 | -1,17 | -0,94 | 3,34 | -1,41 | -0,56 | -0,95 | -0,67 | -2,31 | 10,91 |
| C29 | -0,43 | -2,44 | -1,17 | 6,80 | 0,23 | -1,41 | 14,07 | -0,95 | -0,67 | -2,31 | -0,53 |
| C30 | -0,42 | -2,40 | 1,57 | 19,26 | -3,43 | -1,38 | -0,55 | 3,46 | -0,66 | -2,27 | -0,53 |
| C31 | -0,31 | 0,23 | -0,85 | -0,68 | -0,81 | -1,02 | -0,40 | -0,69 | -0,49 | 3,18 | -0,39 |
| C32 | -0,48 | 3,85 | -1,30 | -1,04 | 1,36 | -1,56 | -0,62 | -1,06 | -0,74 | -2,57 | -0,59 |
| C33 | 12,32 | 2,07 | -0,87 | -0,70 | -1,50 | -1,05 | -0,42 | -0,71 | -0,50 | -0,37 | -0,40 |
| C34 | -0,52 | 3,96 | -1,41 | -1,13 | -0,12 | -1,69 | -0,67 | -1,15 | 4,31 | -1,93 | -0,64 |
| C35 | -0,34 | -0,06 | -0,92 | -0,73 | 2,93 | -1,10 | -0,44 | -0,75 | -0,52 | -1,81 | -0,42 |
| C36 | -0,33 | 1,93 | -0,90 | -0,72 | -2,15 | -1,08 | -0,43 | -0,73 | -0,51 | 3,49 | -0,41 |
| C37 | -0,57 | -3,25 | 6,55 | -1,25 | -0,09 | -1,87 | -0,74 | 11,83 | -0,89 | -3,07 | -0,71 |
| C38 | -0,40 | 13,21 | -1,10 | -0,88 | -5,95 | -1,32 | -0,52 | -0,90 | -0,63 | -2,17 | -0,50 |
| C39 | -0,30 | -1,72 | 8,02 | -0,66 | -1,54 | -0,99 | -0,39 | -0,67 | -0,47 | -0,20 | 5,01 |
| C40 | -0,39 | -2,21 | -1,06 | -0,85 | 2,63 | -1,27 | -0,51 | -0,87 | -0,61 | 1,25 | -0,48 |
| C41 | -0,21 | -1,18 | -0,57 | -0,45 | 2,62 | -0,68 | -0,27 | -0,46 | -0,32 | -1,12 | -0,26 |
| C42 | -0,23 | -1,32 | -0,63 | -0,51 | -0,88 | -0,76 | -0,30 | -0,52 | -0,36 | 4,33 | -0,29 |
| C43 | -0,21 | -1,18 | -0,57 | -0,45 | -0,22 | -0,68 | -0,27 | -0,46 | 12,26 | -1,12 | -0,26 |
| C44 | -0,26 | -1,50 | 6,50 | 8,33 | -3,91 | 2,81 | -0,34 | -0,59 | -0,41 | -1,42 | -0,33 |
| C45 | -0,28 | -1,61 | 1,91 | -0,62 | -4,20 | -0,93 | -0,37 | -0,63 | -0,44 | 8,34 | -0,35 |
| C46 | -0,36 | -2,04 | -0,98 | -0,79 | 2,08 | -1,18 | 6,05 | -0,80 | -0,56 | -0,13 | -0,45 |
| C47 | -0,33 | -1,87 | -0,90 | 5,04 | -1,25 | -1,08 | -0,43 | -0,73 | 3,47 | 2,18 | -0,41 |
| C48 | -0,27 | -1,56 | -0,75 | -0,60 | -4,05 | 6,18 | -0,36 | -0,61 | -0,43 | -1,48 | 23,37 |
| C49 | -0,23 | -1,32 | -0,63 | 5,59 | 1,02 | -0,76 | -0,30 | -0,52 | -0,36 | -1,25 | -0,29 |
| C50 | -0,28 | -1,61 | -0,78 | -0,62 | -0,57 | -0,93 | -0,37 | -0,63 | -0,44 | 4,54 | -0,35 |
| C51 | -0,29 | -1,67 | 17,43 | -0,64 | -3,33 | -0,96 | -0,38 | -0,65 | -0,46 | -1,58 | -0,37 |
| C52 | -0,15 | -0,83 | -0,40 | -0,32 | -0,16 | -0,48 | -0,19 | -0,33 | -0,23 | -0,79 | 10,90 |
| C53 | -0,21 | -1,18 | -0,57 | -0,45 | -0,22 | -0,68 | -0,27 | -0,46 | -0,32 | -1,12 | 15,42 |
| C54 | -0,35 | 8,02 | -0,96 | -0,77 | -2,69 | -1,15 | -0,46 | -0,78 | -0,55 | -1,89 | -0,44 |
| C55 | -0,16 | -0,93 | -0,45 | -0,36 | 2,07 | -0,54 | -0,21 | -0,37 | -0,26 | -0,88 | -0,20 |
| C56 | -0,26 | 0,06 | 9,39 | -0,58 | -1,68 | -0,87 | -0,34 | -0,59 | -0,41 | -1,42 | -0,33 |
| C57 | -0,26 | -1,50 | 18,06 | -0,58 | -3,91 | -0,87 | -0,34 | -0,59 | -0,41 | -1,42 | -0,33 |
| C58 | -0,19 | -1,10 | -0,53 | -0,42 | 2,45 | -0,64 | -0,25 | -0,43 | -0,30 | -1,04 | -0,24 |
| C59 | -0,16 | -0,93 | -0,45 | 5,38 | 0,27 | -0,54 | -0,21 | -0,37 | -0,26 | -0,88 | -0,20 |
| C60 | -0,16 | -0,93 | 4,21 | -0,36 | 0,27 | -0,54 | -0,21 | -0,37 | -0,26 | -0,88 | -0,20 |
| C61 | -0,19 | 3,17 | -0,53 | -0,42 | -0,59 | -0,64 | -0,25 | -0,43 | -0,30 | -1,04 | -0,24 |

**Mkv2_DS**



*Graph of the partition. Node are proportional to the number of agents that belong to the corresponding community. Edges represent overlaps between communities and they are proportional to the number of nodes involved. Fruchterman Reingold layout.*

## a.7. BOOL_1

*Typology of agents (obs. freq)*

| BOOL_1 | GI | TA | CC | SC | LGI | E | R | BS | KIBS |
|---|---|---|---|---|---|---|---|---|---|
| C0 | 1 | 3 | 7 | | 19 | 22 | | | 5 |
| C1 | 2 | | | | | 8 | 1 | | 8 |
| C2 | 2 | | | | 4 | 6 | | 1 | 4 |
| C3 | 2 | 3 | 3 | 1 | 5 | 17 | 7 | | 34 |
| C4 | | 1 | | | | 8 | 1 | | 2 |
| C5 | 1 | | | | 1 | 11 | | 1 | 8 |
| C6 | | 1 | | | 1 | 3 | 4 | | 12 |
| C7 | | | | | 2 | 2 | | | 1 |
| C8 | | 1 | | | 5 | 6 | | 1 | 4 |
| C9 | 1 | 2 | | 1 | 5 | 15 | 4 | 2 | 12 |
| C10 | 4 | | | | 7 | 23 | 7 | 2 | 30 |
| C11 | 4 | | | | 5 | 12 | 4 | 2 | 28 |
| C12 | 3 | | | | 3 | 7 | 1 | 2 | 18 |
| C13 | 1 | | | | 4 | 18 | 5 | 1 | 12 |
| C14 | | | | | | 3 | 3 | | 4 |

*Typology of agents (adj. res)*

| BOOL_1 | GI | TA | CC | SC | LGI | E | R | BS | KIBS |
|---|---|---|---|---|---|---|---|---|---|
| C0 | -0,99 | 1,66 | 5,87 | -0,51 | 5,15 | 1,06 | -2,28 | -1,26 | -4,64 |
| C1 | 1,39 | -0,67 | -0,64 | -0,28 | -1,66 | 0,92 | -0,37 | -0,70 | 0,51 |
| C2 | 1,57 | -0,63 | -0,60 | -0,27 | 1,44 | 0,26 | -1,19 | 0,95 | -1,14 |
| C3 | -0,66 | 1,22 | 1,41 | 1,43 | -1,49 | -1,72 | 0,80 | -1,44 | 2,02 |
| C4 | -0,74 | 1,46 | -0,50 | -0,22 | -1,31 | 2,57 | 0,12 | -0,55 | -1,45 |
| C5 | 0,08 | -0,72 | -0,69 | -0,31 | -1,13 | 1,81 | -1,36 | 0,67 | -0,03 |
| C6 | -0,98 | 0,81 | -0,67 | -0,30 | -1,07 | -1,81 | 2,07 | -0,74 | 2,00 |
| C7 | -0,47 | -0,34 | -0,32 | -0,14 | 1,90 | 0,37 | -0,64 | -0,35 | -0,78 |
| C8 | -0,88 | 1,05 | -0,60 | -0,27 | 2,19 | 0,26 | -1,19 | 0,95 | -1,14 |
| C9 | -0,62 | 1,17 | -0,97 | 2,12 | -0,08 | 0,48 | 0,54 | 1,04 | -1,13 |
| C10 | 0,58 | -1,39 | -1,33 | -0,59 | -0,76 | -0,18 | 0,76 | 0,20 | 0,86 |
| C11 | 1,19 | -1,18 | -1,13 | -0,50 | -0,76 | -1,78 | -0,05 | 0,63 | 2,33 |
| C12 | 1,38 | -0,91 | -0,87 | -0,38 | -0,64 | -1,52 | -1,04 | 1,37 | 2,05 |
| C13 | -0,59 | -1,01 | -0,96 | -0,43 | -0,51 | 1,64 | 1,21 | 0,01 | -1,02 |
| C14 | -0,67 | -0,48 | -0,46 | -0,20 | -1,20 | -0,16 | 2,75 | -0,50 | 0,22 |

*Participations of agents in waves (obs. freq)*

| BOOL_1 | 2002_ 171 | 2002_ 172 | 2002_ ITT | 2004_ 171 | 2004_ 171A | 2005_ 171 | 2006_ VIN | 2007_ 171 | 2008_ 171 |
|---|---|---|---|---|---|---|---|---|---|
| C0 | | | | 12 | 19 | 109 | | 14 | |
| C1 | | | | | | 44 | | | |
| C2 | | | | | 3 | 51 | | | |
| C3 | 36 | 37 | 84 | 16 | 8 | 109 | | 11 | |
| C4 | | | | | | | | 26 | |
| C5 | 10 | | 8 | 15 | 7 | 12 | 16 | | |
| C6 | 7 | 10 | 9 | | 9 | 23 | | 4 | |
| C7 | | | | | | 11 | | | |
| C8 | | | | | | 46 | | | |
| C9 | 33 | 11 | 26 | 7 | 2 | 52 | | | 2 |
| C10 | 12 | 36 | 112 | 14 | 13 | 66 | 12 | 11 | |
| C11 | 8 | 27 | 85 | 12 | 13 | 57 | 12 | 9 | |
| C12 | | 23 | 40 | 8 | 4 | 37 | 8 | 5 | |
| C13 | 12 | 13 | 45 | 9 | 15 | 35 | 4 | 10 | |
| C14 | 19 | | 73 | | | | | | |

*Participations of agents in waves (adj. res.)*

| BOOL_1 | 2002_ 171 | 2002_ 172 | 2002_ ITT | 2004_ 171 | 2004_ 171A | 2005_ 171 | 2006_ VIN | 2007_ 171 | 2008_ 171 |
|---|---|---|---|---|---|---|---|---|---|
| C0 | -3,78 | -4,07 | -7,99 | 1,45 | 4,09 | 9,06 | -2,27 | 2,34 | -0,44 |
| C1 | -1,95 | -2,10 | -4,13 | -1,59 | -1,59 | 8,75 | -1,17 | -1,56 | -0,23 |
| C2 | -2,17 | -2,34 | -4,59 | -1,76 | 0,09 | 8,86 | -1,30 | -1,73 | -0,25 |
| C3 | 2,96 | 2,25 | 0,21 | 0,02 | -2,24 | -0,35 | -3,33 | -1,27 | -0,64 |
| C4 | -1,49 | -1,61 | -3,16 | -1,21 | -1,21 | -3,94 | -0,90 | 22,12 | -0,17 |
| C5 | 2,17 | -2,63 | -2,95 | 6,30 | 1,88 | -3,39 | 10,21 | -1,95 | -0,28 |
| C6 | 1,05 | 2,02 | -2,32 | -1,90 | 3,30 | | -1,40 | 0,49 | -0,27 |
| C7 | -0,97 | -1,04 | -2,05 | -0,79 | -0,79 | 4,33 | -0,58 | -0,77 | -0,11 |
| C8 | -2,00 | -2,15 | -4,22 | -1,62 | -1,62 | 8,95 | -1,20 | -1,60 | -0,23 |
| C9 | 7,62 | -0,28 | -2,12 | -0,01 | -2,03 | 0,50 | -2,09 | -2,79 | 4,95 |
| C10 | -2,33 | 2,61 | 5,34 | -0,18 | -0,47 | -4,94 | 1,48 | -0,93 | -0,61 |
| C11 | -2,51 | 1,78 | 3,83 | 0,07 | 0,39 | -3,81 | 2,29 | -0,79 | -0,54 |
| C12 | -3,37 | 3,85 | 1,19 | 0,58 | -1,08 | -1,80 | 2,36 | -0,59 | -0,39 |
| C13 | 0,28 | 0,07 | 1,13 | 0,56 | 2,90 | -3,26 | -0,12 | 1,06 | -0,42 |
| C14 | 4,73 | -3,09 | 11,47 | -2,33 | -2,33 | -7,57 | -1,72 | -2,29 | -0,33 |

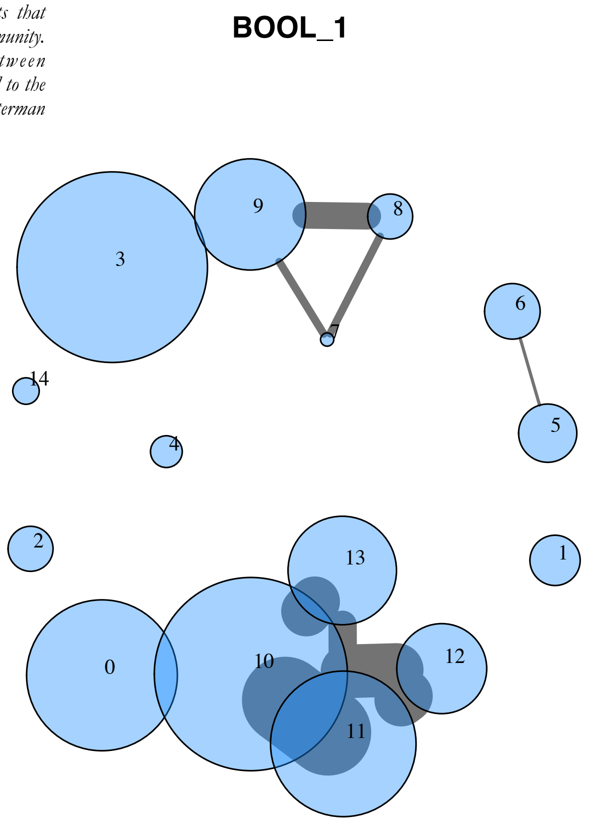
*Participations of agents in projects with specific technological domains (obs. freq)*

| BOOL_1 | Others | Bio-Technologies | Organic Chemistry | Geotermal Sciences and Biomasses | ICT - Multimedia | Mechanics Engineering | Multi Disciplinary | Nanotechnologies | New Materials | Optoelectronic | NA |
|---|---|---|---|---|---|---|---|---|---|---|---|
| C0 | | | 2 | 15 | 81 | 30 | | | 4 | 16 | 6 |
| C1 | | 4 | | | 38 | | 2 | | | 0 | 0 |
| C2 | | | | | 39 | 6 | | | 9 | 0 | 0 |
| C3 | 4 | 24 | 31 | | 174 | 16 | | 4 | 8 | 38 | 2 |
| C4 | 4 | | | | 4 | 8 | | | | 2 | 8 |
| C5 | | 6 | 7 | 5 | 28 | 18 | | | | 4 | 0 |
| C6 | | 4 | 7 | | 30 | 7 | 5 | 4 | 2 | 3 | 0 |
| C7 | | 2 | | | 9 | | | | | 0 | 0 |
| C8 | | 5 | | | 32 | 7 | | | 2 | 0 | 0 |
| C9 | | 5 | 7 | 7 | 85 | 7 | | 16 | 2 | 2 | 2 |
| C10 | 8 | 33 | 9 | 10 | 133 | | 6 | 9 | 8 | 58 | 2 |
| C11 | 4 | 23 | 9 | 10 | 113 | | 6 | 9 | 8 | 41 | 0 |
| C12 | 4 | 6 | | 10 | 70 | | 3 | | 8 | 24 | 0 |
| C13 | 6 | 6 | 9 | 3 | 70 | | | 15 | | 30 | 4 |
| C14 | | 38 | 2 | | 28 | | | | | 24 | 0 |

*Participations of agents in projects with specific technological domains (adj. res.)*

| BOOL_1 | Others | Bio-Technologies | Organic Chemistry | Geotermal Sciences and Biomasses | ICT - Multimedia | Mechanics Engineering | Multi Disciplinary | Nanotechnologies | New Materials | Optoelectronic | NA |
|---|---|---|---|---|---|---|---|---|---|---|---|
| C0 | -1,71 | -4,05 | -2,10 | 4,53 | -0,14 | 7,81 | -1,46 | -2,38 | -0,24 | -1,27 | 2,83 |
| C1 | -0,89 | 0,05 | -1,50 | -1,26 | 4,47 | -1,64 | 1,99 | -1,23 | -1,16 | -2,68 | -0,79 |
| C2 | -0,98 | -2,33 | -1,66 | -1,40 | 2,86 | 1,77 | -0,84 | -1,37 | 6,12 | -2,98 | -0,88 |
| C3 | -0,56 | -0,60 | 5,01 | -3,58 | 1,79 | -0,26 | -2,15 | -2,06 | -0,28 | -0,63 | -1,15 |
| C4 | 5,43 | -1,60 | -1,14 | -0,97 | -3,89 | 5,60 | -0,58 | -0,94 | -0,89 | -0,91 | 13,02 |
| C5 | -1,11 | -0,02 | 2,21 | 1,83 | -2,01 | 7,60 | -0,95 | -1,54 | -1,45 | -1,92 | -0,99 |
| C6 | -1,06 | -0,68 | 2,48 | -1,51 | -0,76 | 1,97 | 4,91 | 1,45 | 0,16 | -2,08 | -0,94 |
| C7 | -0,44 | 1,09 | -0,74 | -0,63 | 1,91 | -0,81 | -0,38 | -0,61 | -0,58 | -1,33 | -0,39 |
| C8 | -0,91 | 0,48 | -1,53 | -1,29 | 2,26 | 2,86 | -0,77 | -1,26 | 0,59 | -2,75 | -0,81 |
| C9 | -1,58 | -2,16 | 0,31 | 1,22 | 2,59 | -0,19 | -1,35 | 5,95 | -1,00 | -4,27 | 0,14 |
| C10 | 1,67 | 1,96 | -1,25 | 0,21 | -1,79 | -4,42 | 1,50 | 0,02 | | 3,81 | -1,00 |
| C11 | 0,11 | 0,81 | -0,52 | 0,94 | -0,79 | -3,90 | 2,07 | 0,72 | 0,65 | 2,14 | -1,88 |
| C12 | 1,34 | -1,66 | -2,58 | 2,93 | 0,67 | -2,83 | 1,20 | -2,12 | 2,42 | 1,83 | -1,37 |
| C13 | 2,40 | -2,05 | 0,93 | -0,90 | -1,04 | -3,05 | -1,41 | 5,11 | -2,16 | 2,61 | 1,54 |
| C14 | -1,30 | 11,24 | -1,18 | -1,85 | -4,48 | -2,41 | -1,11 | -1,80 | -1,70 | 3,52 | -1,16 |

*Graph of the partition. Node are proportional to the number of agents that belong to the corresponding community. Edges represent overlaps between communities and they are proportional to the number of nodes involved. Fruchterman Reingold layout.*



## a.8. BOOL_2

*Typology of agents (obs. freq)*

| BOOL_2 | GI | TA | CC | SC | LGI | E | R | BS | KIBS |
|---|---|---|---|---|---|---|---|---|---|
| C0 | | 1 | 1 | | 2 | 9 | 2 | | 7 |
| C1 | | 1 | | | | 7 | | | 3 |
| C2 | | | 7 | | 2 | 2 | 2 | | 8 |
| C3 | 1 | 4 | 4 | 1 | 5 | 22 | 6 | | 27 |
| C4 | | 2 | | | 2 | 6 | 1 | | 7 |
| C5 | 1 | 2 | | | 2 | 11 | 1 | | 4 |
| C6 | 2 | 3 | 4 | 1 | 4 | 23 | 7 | 1 | 26 |
| C7 | | 2 | | | 5 | 21 | 1 | 1 | 13 |
| C8 | 1 | | | | 2 | 3 | 2 | 1 | 5 |
| C9 | 2 | 4 | | 1 | 7 | 19 | 7 | | 29 |
| C10 | | 2 | 10 | | 5 | 23 | 12 | 1 | 29 |
| C11 | 2 | 1 | | 2 | 2 | 16 | 2 | | 9 |
| C12 | 2 | 3 | | 2 | 4 | 26 | 4 | 1 | 15 |

*Typology of agents (adj. res)*

| BOOL_2 | GI | TA | CC | SC | LGI | E | R | BS | KIBS |
|---|---|---|---|---|---|---|---|---|---|
| C0 | -0,70 | -0,03 | -0,07 | -0,55 | 0,22 | 0,57 | 0,05 | -0,47 | -0,24 |
| C1 | -0,49 | 0,70 | -0,76 | -0,39 | -0,98 | 1,99 | -1,04 | -0,33 | -0,49 |
| C2 | -0,68 | -1,04 | 6,18 | -0,54 | 0,29 | -2,52 | 0,12 | -0,46 | 0,39 |
| C3 | -0,40 | 0,44 | 0,35 | 0,09 | -0,25 | -0,72 | -0,08 | -0,87 | 0,84 |
| C4 | -0,63 | 1,31 | -0,98 | -0,50 | 0,52 | -0,18 | -0,50 | -0,42 | 0,43 |
| C5 | 0,89 | 1,07 | -1,06 | -0,54 | 0,29 | 1,67 | -0,67 | -0,46 | -1,49 |
| C6 | 0,48 | -0,20 | 0,32 | 0,08 | -0,76 | -0,55 | 0,33 | 0,44 | 0,47 |
| C7 | -0,99 | -0,01 | -1,55 | -0,79 | 0,95 | 1,94 | -1,57 | 0,98 | -0,56 |
| C8 | 1,36 | -0,84 | -0,86 | -0,44 | 0,90 | -1,10 | 0,73 | 2,44 | 0,13 |
| C9 | 0,52 | 0,47 | -2,02 | 0,11 | 0,75 | -1,44 | 0,42 | -0,87 | 1,48 |
| C10 | -1,43 | -1,05 | 3,34 | -1,14 | -0,65 | -1,49 | 2,02 | 0,29 | 0,25 |
| C11 | 1,62 | -0,50 | -1,37 | 2,42 | -0,45 | 1,49 | -0,62 | -0,59 | -0,98 |
| C12 | 0,81 | 0,22 | -1,81 | 1,54 | -0,26 | 1,73 | -0,51 | 0,68 | -1,32 |

*Participations of agents in waves (obs. freq)*

| BOOL_2 | 2002_ 171 | 2002_ 172 | 2002_ ITT | 2004_ 171 | 2004_ 171A | 2005_ 171 | 2006_ VIN | 2007_ 171 | 2008_ 171 |
|---|---|---|---|---|---|---|---|---|---|
| C0 | | | | | 4 | 45 | | 23 | |
| C1 | | | | | | | | 23 | |
| C2 | | | | 8 | 12 | 94 | | | |
| C3 | 45 | 34 | 124 | 16 | 12 | 90 | | 16 | |
| C4 | 9 | 17 | 11 | | 3 | 3 | | 4 | |
| C5 | 2 | 9 | 25 | | | 17 | | | |
| C6 | 41 | 29 | 81 | 19 | 14 | 90 | | 27 | |
| C7 | | | | 2 | 2 | 69 | 6 | 26 | 2 |
| C8 | 19 | 8 | 2 | 5 | | 13 | | | |
| C9 | 12 | 21 | 76 | | 20 | 70 | | 4 | |
| C10 | 41 | 37 | 126 | 34 | 12 | 125 | | 27 | |
| C11 | 14 | 10 | 50 | 2 | 2 | 15 | | | 2 |
| C12 | 19 | 15 | 81 | 2 | 2 | 34 | | | 2 |

*Participations of agents in waves (adj. res.)*

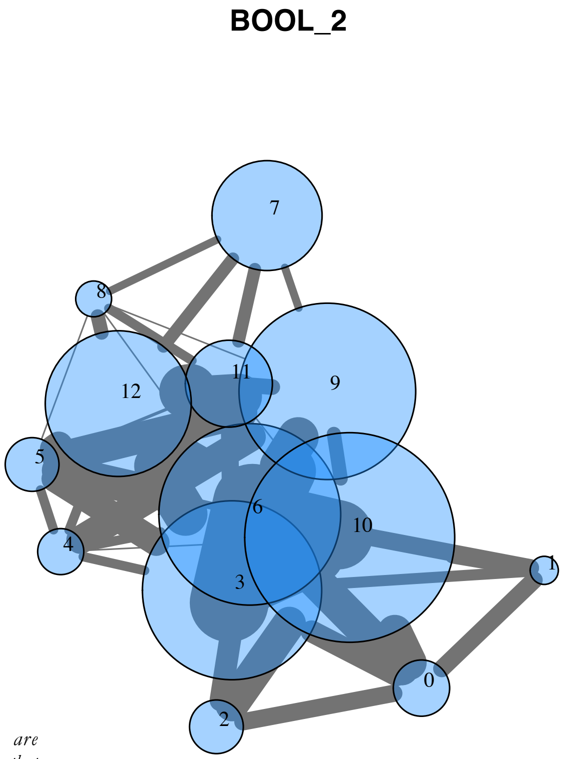
| BOOL_2 | 2002_ 171 | 2002_ 172 | 2002_ ITT | 2004_ 171 | 2004_ 171A | 2005_ 171 | 2006_ VIN | 2007_ 171 | 2008_ 171 |
|---|---|---|---|---|---|---|---|---|---|
| C0 | -2,93 | -2,75 | -5,59 | -1,88 | 0,56 | 5,20 | -0,48 | 7,89 | -0,48 |
| C1 | -1,64 | -1,54 | -3,12 | -1,05 | -1,02 | -3,46 | -0,27 | 16,74 | -0,27 |
| C2 | -3,73 | -3,50 | -7,11 | 1,34 | 3,43 | 11,26 | -0,61 | -3,17 | -0,61 |
| C3 | 2,01 | 0,62 | 3,25 | 0,24 | -0,68 | -3,11 | -1,12 | -2,22 | -1,12 |
| C4 | 2,01 | 6,47 | -0,92 | -1,51 | 0,74 | -4,05 | -0,39 | 0,22 | -0,39 |
| C5 | -1,59 | 1,99 | 2,87 | -1,60 | -1,55 | -0,30 | -0,41 | -2,13 | -0,41 |
| C6 | 2,04 | 0,28 | -1,05 | 1,65 | 0,38 | -1,63 | -1,05 | 0,92 | -1,05 |
| C7 | -3,61 | -3,39 | -6,87 | -1,35 | -1,25 | 6,85 | 10,20 | 6,65 | 3,01 |
| C8 | 6,86 | 1,88 | -3,84 | 2,06 | -1,46 | -0,93 | -0,39 | -2,00 | -0,39 |
| C9 | -2,18 | 0,60 | 2,64 | -3,27 | 4,19 | 0,15 | -0,84 | -3,22 | -0,84 |
| C10 | -0,10 | | 0,94 | 4,30 | -1,40 | -1,38 | -1,25 | -0,81 | -1,25 |
| C11 | 1,45 | 0,46 | 5,08 | -1,15 | -1,06 | -3,84 | -0,55 | -2,88 | 3,25 |
| C12 | 0,82 | 0,21 | 6,49 | -2,01 | -1,90 | -3,30 | -0,72 | -3,74 | 2,31 |

*Participations of agents in projects with specific technological domains (obs. freq)*

| BOOL_2 | Others | Bio-Technologies | Organic Chemistry | Geotermal Sciences and Biomasses | ICT - Multimedia | Mechanics Engineering | Multi Disciplinary | Nanotechnologies | New Materials | Optoelectronic | NA |
|---|---|---|---|---|---|---|---|---|---|---|---|
| C0 | | | 3 | 4 | 37 | 8 | | | 4 | 12 | 4 |
| C1 | | | 3 | | 10 | | | | | 6 | 4 |
| C2 | | | 16 | 4 | 59 | 16 | 3 | | 4 | 12 | |
| C3 | 4 | 40 | 30 | 4 | 172 | 8 | | 4 | 8 | 61 | 6 |
| C4 | 2 | 8 | 7 | | 27 | | | | | 3 | |
| C5 | 4 | 17 | 4 | | 24 | | | 2 | | 2 | |
| C6 | 4 | 32 | 33 | 7 | 155 | 14 | | 2 | 8 | 38 | 8 |
| C7 | 4 | 5 | 2 | 2 | 65 | 15 | | | 2 | 2 | 10 |
| C8 | | 2 | 5 | 5 | 35 | | | | | | |
| C9 | 4 | 23 | 13 | | 99 | 9 | 5 | 17 | 2 | 31 | |
| C10 | 4 | 63 | 29 | 7 | 186 | 40 | 3 | 2 | 8 | 54 | 6 |
| C11 | 6 | 2 | 2 | 2 | 49 | | | 18 | | 14 | 2 |
| C12 | 6 | 23 | 4 | 2 | 82 | | | 20 | | 16 | 2 |

*Participations of agents in projects with specific technological domains (adj. res.)*

| BOOL_2 | Others | Bio-Technologies | Organic Chemistry | Geotermal Sciences and Biomasses | ICT - Multimedia | Mechanics Engineering | Multi Disciplinary | Nanotechnologies | New Materials | Optoelectronic | NA |
|---|---|---|---|---|---|---|---|---|---|---|---|
| C0 | -1,22 | -3,04 | -1,15 | 2,33 | 0,05 | 2,06 | -0,65 | -1,60 | 2,39 | 0,99 | 2,03 |
| C1 | -0,68 | -1,70 | 0,96 | -0,67 | -0,74 | -1,18 | -0,36 | -0,89 | -0,66 | 1,91 | 5,07 |
| C2 | -1,55 | -3,87 | 2,60 | 1,31 | 0,14 | 4,02 | 3,04 | -2,04 | 1,37 | -0,76 | -1,63 |
| C3 | -1,11 | 0,57 | 0,89 | -1,04 | -0,04 | -2,85 | -1,52 | -2,41 | 0,80 | 3,18 | -0,51 |
| C4 | 1,16 | 1,34 | 1,87 | -0,96 | 0,88 | -1,69 | -0,52 | -1,29 | -0,95 | -1,34 | -1,03 |
| C5 | 3,00 | 4,98 | -0,05 | -1,03 | -0,86 | -1,80 | -0,56 | 0,19 | -1,01 | -2,00 | -1,09 |
| C6 | -0,84 | -0,22 | 2,29 | 0,60 | 0,14 | -0,80 | -1,42 | -2,80 | 1,15 | -0,12 | 0,66 |
| C7 | 1,38 | -2,15 | -2,33 | -0,02 | 2,05 | 3,88 | -0,80 | -1,97 | 0,02 | -3,49 | 5,28 |
| C8 | -0,98 | -1,50 | 0,76 | 4,46 | 3,24 | -1,69 | -0,52 | -1,29 | -0,95 | -2,66 | -1,03 |
| C9 | 0,03 | 0,16 | -0,74 | -2,09 | -0,71 | -0,78 | 3,83 | 4,24 | -0,96 | 1,10 | -2,23 |
| C10 | -1,55 | 3,37 | -0,43 | -0,25 | -2,19 | 4,22 | 0,55 | -3,55 | 0,25 | 0,40 | -1,02 |
| C11 | 3,17 | -2,84 | -2,10 | 0,16 | 0,09 | -2,44 | -0,75 | 8,71 | -1,37 | 0,57 | -0,03 |
| C12 | 1,81 | 1,60 | -2,50 | -0,57 | 0,46 | -3,17 | -0,98 | 6,94 | -1,78 | -0,97 | -0,77 |




*Graph of the partition. Node are proportional to the number of agents that belong to the corresponding community. Edges represent overlaps between communities and they are proportional to the number of nodes involved. Fruchterman Reingold layout.*

## a.9. ∂LoA

*Typology of agents (obs. freq)*

| ∂LoA | GI | TA | CC | SC | LGI | E | R | BS | KIBS |
|---|---|---|---|---|---|---|---|---|---|
| C0 | | | | | 2 | | | 1 | 1 |
| C1 | | 1 | | | 2 | | | 1 | 1 |
| C2 | | | | | 2 | 5 | | 1 | 2 |
| C3 | 2 | | | 1 | 2 | 12 | 3 | 1 | 8 |
| C4 | 2 | 2 | | 1 | 4 | 15 | 4 | | 11 |
| C5 | 1 | | 6 | | 2 | 4 | 3 | 1 | 8 |
| C6 | 1 | 4 | | | 8 | 17 | 5 | 1 | 12 |
| C7 | 2 | 3 | 2 | 1 | 11 | 25 | 8 | 2 | 35 |
| C8 | 2 | 1 | | | 9 | 24 | 2 | 1 | 11 |
| C9 | 3 | 2 | | 1 | 6 | 25 | 7 | 1 | 18 |
| C10 | 2 | 1 | 8 | | 9 | 22 | 8 | 2 | 38 |
| C11 | 2 | | 2 | 1 | 6 | 20 | 14 | 5 | 40 |
| C12 | 2 | | 2 | 1 | 10 | 19 | 13 | 1 | 42 |
| C13 | 1 | | 5 | 1 | 4 | 23 | 14 | 2 | 39 |
| C14 | 3 | | 5 | 1 | 5 | 19 | 15 | 1 | 41 |
| C15 | 2 | | 1 | 1 | 9 | 16 | 13 | 1 | 43 |

*Typology of agents (adj. res)*

| ∂LoA | GI | TA | CC | SC | LGI | E | R | BS | KIBS |
|---|---|---|---|---|---|---|---|---|---|
| C0 | -0,34 | -0,25 | -0,38 | -0,20 | 2,65 | -1,23 | -0,75 | 2,92 | -0,58 |
| C1 | -0,38 | 3,34 | -0,42 | -0,23 | 2,22 | -1,38 | -0,83 | 2,54 | -0,87 |
| C2 | -0,54 | -0,40 | -0,60 | -0,32 | 1,04 | 1,61 | -1,18 | 1,55 | -1,24 |
| C3 | 1,37 | -0,69 | -1,04 | 1,34 | -0,59 | 1,71 | -0,30 | 0,35 | -1,28 |
| C4 | 0,91 | 1,84 | -1,21 | 1,00 | 0,02 | 1,58 | -0,37 | -1,01 | -1,42 |
| C5 | 0,37 | -0,64 | 5,70 | -0,51 | -0,36 | -1,30 | -0,02 | 0,51 | -0,73 |
| C6 | -0,30 | 3,89 | -1,35 | -0,72 | 1,54 | 1,28 | -0,38 | -0,17 | -2,05 |
| C7 | -0,33 | 1,45 | -0,66 | 0,12 | 0,73 | 0,15 | -0,96 | -0,13 | 0,06 |
| C8 | 0,54 | 0,26 | -1,38 | -0,73 | 1,89 | 3,36 | -1,82 | -0,21 | -2,54 |
| C9 | 0,99 | 1,07 | -1,56 | 0,48 | -0,17 | 2,26 | -0,26 | -0,46 | -1,76 |
| C10 | -0,34 | -0,36 | 2,98 | -1,01 | -0,05 | -0,67 | -1,00 | -0,15 | 0,66 |
| C11 | -0,34 | -1,26 | -0,68 | 0,11 | -1,15 | -1,17 | 1,04 | 2,01 | 1,11 |
| C12 | -0,34 | -1,26 | -0,68 | 0,11 | 0,32 | -1,42 | 0,70 | -0,87 | 1,57 |
| C13 | -1,01 | -1,25 | 1,18 | 0,12 | -1,86 | -0,35 | 1,09 | -0,13 | 0,98 |
| C14 | 0,33 | -1,26 | 1,15 | 0,11 | -1,52 | -1,42 | 1,38 | -0,87 | 1,34 |
| C15 | -0,27 | -1,23 | -1,22 | 0,16 | 0,10 | -1,93 | 0,89 | -0,81 | 2,20 |

*Participations of agents in waves (obs. freq)*

| ∂LoA | 2002_ 171 | 2002_ 172 | 2002_ ITT | 2004_ 171 | 2004_ 171A | 2005_ 171 | 2006_ VIN | 2007_ 171 | 2008_ 171 | |
|---|---|---|---|---|---|---|---|---|---|---|
| C0 | | | | | | 17 | | | | |
| C1 | | | | | 3 | 17 | | | | |
| C2 | | | | | | 29 | | | | |
| C3 | 10 | 8 | 42 | | | 15 | | | 2 | |
| C4 | 2 | 11 | 62 | | | 31 | | | | |
| C5 | | | | 11 | 12 | 66 | | 11 | | 2 |
| C6 | 9 | 21 | 42 | | 19 | 48 | | 2 | | |
| C7 | 38 | 29 | 89 | 5 | 29 | 88 | | 15 | 2 | |
| C8 | | | | 19 | 4 | 95 | 6 | 16 | 2 | |
| C9 | 29 | 19 | 83 | 2 | 2 | 46 | | | 2 | 2 |
| C10 | 45 | 27 | 59 | 22 | 39 | 110 | | 21 | | 2 |
| C11 | 54 | 56 | 164 | 40 | 30 | 111 | 24 | | | 2 |
| C12 | 85 | 54 | 160 | 44 | 34 | 106 | 16 | 8 | | |
| C13 | 64 | 57 | 189 | 40 | 30 | 115 | 17 | 4 | | |
| C14 | 68 | 53 | 187 | 32 | 25 | 110 | 20 | 2 | | |
| C15 | 87 | 53 | 128 | 36 | 27 | 114 | 12 | 9 | | |

*Participations of agents in waves (adj. res.)*

| ∂LoA | 2002_ 171 | 2002_ 172 | 2002_ ITT | 2004_ 171 | 2004_ 171A | 2005_ 171 | 2006_ VIN | 2007_ 171 | 2008_ 171 |
|---|---|---|---|---|---|---|---|---|---|
| C0 | -1,57 | -1,37 | -2,76 | -1,08 | -1,09 | 6,52 | -0,65 | -0,63 | -0,19 |
| C1 | -1,70 | -1,49 | -3,00 | -1,18 | 1,54 | 5,58 | -0,71 | -0,68 | -0,20 |
| C2 | -2,05 | -1,80 | -3,62 | -1,42 | -1,43 | 8,52 | -0,85 | -0,82 | -0,25 |
| C3 | 0,10 | 0,13 | 4,53 | -2,33 | -2,34 | -1,80 | -1,40 | -1,35 | 4,69 |
| C4 | -3,37 | 0,15 | 6,23 | -2,74 | -2,76 | 0,13 | -1,65 | -1,59 | -0,47 |
| C5 | -3,85 | -3,37 | -6,78 | 1,88 | 2,25 | 8,36 | -1,60 | 5,96 | -0,46 |
| C6 | -2,27 | 2,00 | -0,30 | -3,17 | 3,41 | 1,43 | -1,91 | -0,68 | -0,55 |
| C7 | 0,15 | -0,07 | -0,29 | -3,45 | 2,40 | 0,45 | -2,82 | 3,40 | 1,87 |
| C8 | -4,61 | -4,04 | -8,12 | 3,43 | -1,82 | 10,26 | 1,41 | 7,36 | 3,23 |
| C9 | 1,36 | 0,20 | 4,33 | -3,02 | -3,05 | -1,09 | -2,19 | -2,11 | 2,72 |
| C10 | 0,76 | -1,00 | -5,14 | 0,28 | 4,23 | 2,23 | -2,97 | 5,36 | -0,85 |
| C11 | -0,93 | 1,36 | 1,68 | 1,82 | -0,24 | -2,85 | 3,90 | -3,55 | -1,06 |
| C12 | 3,03 | 0,56 | 0,34 | 2,20 | 0,19 | -4,15 | 1,13 | -1,11 | -1,10 |
| C13 | -0,14 | 0,89 | 3,02 | 1,30 | -0,69 | -3,45 | 1,36 | -2,43 | -1,11 |
| C14 | 0,78 | 0,57 | 3,47 | 0,00 | -1,44 | -3,46 | 2,46 | -2,98 | -1,08 |
| C15 | 4,21 | 1,09 | -1,72 | 1,21 | -0,67 | -2,15 | 0,21 | -0,51 | -1,04 |

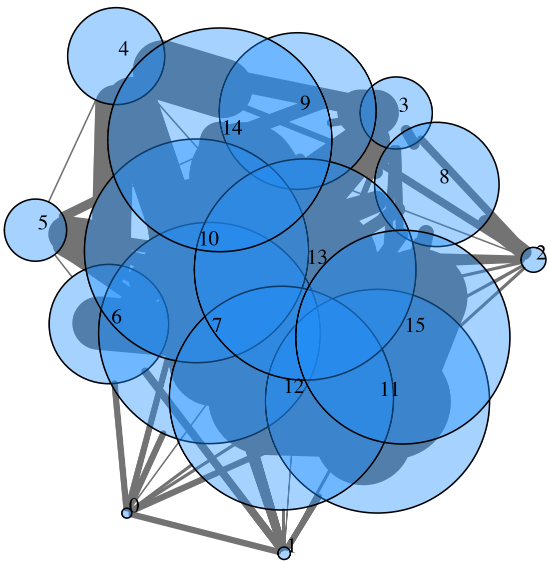
*Participations of agents in projects with specific technological domains (obs. freq)*

| ∂LoA | Others | Bio-Technologies | Organic Chemistry | Geotermal Sciences and Biomasses | ICT - Multimedia | Mechanics Engineering | Multi Disciplinary | Nanotechnologies | New Materials | Optoelectronic | NA |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **C0** | -0,49 | -1,39 | -1,15 | -0,76 | 4,08 | -0,90 | -0,30 | -1,07 | -0,53 | -1,66 | -0,31 |
| **C1** | -0,53 | -1,50 | 1,35 | -0,83 | 3,08 | -0,98 | -0,33 | -1,16 | -0,57 | -1,80 | -0,34 |
| **C2** | -0,64 | 1,28 | -1,51 | -1,00 | 0,86 | 5,07 | -0,40 | -1,39 | -0,69 | -2,17 | -0,41 |
| **C3** | 2,89 | -1,06 | -2,47 | -1,63 | 1,84 | -1,94 | -0,65 | 2,95 | -1,13 | -0,89 | 2,41 |
| **C4** | 2,13 | 1,73 | -0,63 | -1,92 | -0,73 | -1,34 | -0,77 | -1,06 | -1,33 | 2,66 | -0,79 |
| **C5** | -1,20 | -3,40 | 4,62 | 2,11 | -1,55 | 5,55 | 3,41 | -2,61 | 1,95 | -2,30 | 1,94 |
| **C6** | 1,50 | 3,63 | 0,94 | -1,27 | -0,25 | -1,82 | -0,89 | 0,78 | 0,52 | -2,62 | -0,91 |
| **C7** | 0,99 | -2,78 | -0,53 | -0,91 | 0,43 | -1,01 | -1,32 | 2,91 | 1,12 | 0,02 | 3,50 |
| **C8** | 2,95 | -2,65 | -1,40 | 4,96 | 2,92 | 0,21 | -0,89 | -2,42 | 1,19 | -4,37 | 5,93 |
| **C9** | 2,25 | 0,38 | -0,64 | -1,70 | 1,25 | -2,31 | -1,02 | 1,13 | -1,76 | -0,30 | 0,98 |
| **C10** | -1,23 | -5,52 | 2,63 | 1,11 | 0,27 | 2,58 | 1,00 | 2,85 | 2,72 | -2,82 | 1,69 |
| **C11** | -1,10 | 0,58 | 1,22 | 0,07 | -1,73 | -1,61 | 0,28 | -0,78 | 0,93 | 2,92 | -1,76 |
| **C12** | -1,23 | -0,36 | 0,08 | 0,90 | 0,11 | -0,72 | 0,18 | -0,32 | -1,55 | 1,62 | -1,82 |
| **C13** | -1,27 | 2,96 | 0,33 | -0,25 | -2,88 | -0,58 | 0,14 | -1,21 | -0,08 | 3,49 | -1,84 |
| **C14** | -1,19 | 3,95 | -2,38 | -0,09 | -1,04 | 2,14 | -1,76 | -0,99 | -0,35 | 0,58 | -1,80 |
| **C15** | -0,19 | 0,31 | -0,88 | -0,64 | 1,37 | -0,30 | 2,35 | 0,19 | -1,35 | -0,79 | -1,73 |

*Participations of agents in projects with specific technological domains (adj. res.)*

| ∂LoA | Others | Bio-Technologies | Organic Chemistry | Geotermal Sciences and Biomasses | ICT - Multimedia | Mechanics Engineering | Multi Disciplinary | Nanotechnologies | New Materials | Optoelectronic | NA |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **C0** | | | | | 17 | | | | | | |
| **C1** | | | 3 | | 17 | | | | | | |
| **C2** | | 5 | | | 17 | 7 | | | | | |
| **C3** | 4 | 5 | | | 47 | | | 11 | | 8 | 2 |
| **C4** | 4 | 16 | 6 | | 50 | 2 | | 4 | | 24 | |
| **C5** | | | 19 | 7 | 43 | 16 | 3 | | 4 | 6 | 2 |
| **C6** | 4 | 27 | 13 | 2 | 70 | 2 | | 11 | 3 | 9 | |
| **C7** | 6 | 16 | 19 | 7 | 153 | 10 | | 30 | 7 | 41 | 6 |
| **C8** | 6 | 5 | 6 | 15 | 89 | 7 | | 2 | 4 | 2 | 6 |
| **C9** | 6 | 20 | 11 | 2 | 101 | 2 | | 15 | | 24 | 2 |
| **C10** | 2 | 4 | 35 | 14 | 166 | 24 | 3 | 32 | 11 | 28 | 4 |
| **C11** | 4 | 52 | 41 | 16 | 225 | 15 | 3 | 26 | 10 | 87 | |
| **C12** | 4 | 49 | 37 | 20 | 258 | 20 | 3 | 30 | 4 | 82 | |
| **C13** | 4 | 71 | 39 | 16 | 231 | 21 | 3 | 26 | 8 | 97 | |
| **C14** | 4 | 75 | 23 | 16 | 241 | 32 | | 26 | 7 | 73 | |
| **C15** | 6 | 49 | 29 | 13 | 250 | 20 | 6 | 30 | 4 | 59 | |

**dLoA**




*Graph of the partition. Node are proportional to the number of agents that belong to the corresponding community. Edges represent overlaps between communities and they are proportional to the number of nodes involved. Fruchterman Reingold layout.*

## b. Results of the probit models

### b.1. Probit model with 'd_net1' as dependent variable

```
. probit  d_net1          N_int_abvTHR jacc_max dim_int_with_maxJacc   d_R d_KIBS, robust

Iteration 0:   log pseudolikelihood = -242,87299
Iteration 1:   log pseudolikelihood = -236,80581
Iteration 2:   log pseudolikelihood = -236,80114
Iteration 3:   log pseudolikelihood = -236,80114

Probit regression                                 Number of obs   =        352
                                                  Wald chi2(5)    =      11,77
                                                  Prob > chi2     =     0,0381
Log pseudolikelihood = -236,80114                 Pseudo R2       =     0,0250

------------------------------------------------------------------------------------
                     |               Robust
              d_net1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------+--------------------------------------------------------------
        N_int_abvTHR |   ,0047933   ,0032207     1,49   0,137    -,0015192    ,0111058
            jacc_max |  -,0196872    ,012609    -1,56   0,118    -,0444004     ,005026
dim_int_with_maxJacc |   ,0148255   ,0119585     1,24   0,215    -,0086128    ,0382638
                 d_R |   ,0531119   ,2629306     0,20   0,840    -,4622225    ,5684463
              d_KIBS |   ,3023922   ,1498151     2,02   0,044     ,0087601    ,5960244
               _cons |   -,237279   ,1434109    -1,65   0,098    -,5183592    ,0438013
------------------------------------------------------------------------------------
```

### b.2. Probit model with 'd_net2' as dependent variable

```
. probit d_net2           N_int_abvTHR jacc_max dim_int_with_maxJacc   d_R d_KIBS, robust

Iteration 0:   log pseudolikelihood =  -134,5561
Iteration 1:   log pseudolikelihood = -100,79229
Iteration 2:   log pseudolikelihood = -99,286516
Iteration 3:   log pseudolikelihood = -99,279282
Iteration 4:   log pseudolikelihood = -99,279282

Probit regression                                 Number of obs   =        352
                                                  Wald chi2(5)    =      54,36
                                                  Prob > chi2     =     0,0000
Log pseudolikelihood = -99,279282                 Pseudo R2       =     0,2622

------------------------------------------------------------------------------------
                     |               Robust
              d_net2 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------+--------------------------------------------------------------
        N_int_abvTHR |   ,0133035   ,0038935     3,42   0,001     ,0056724    ,0209347
            jacc_max |   ,0217824   ,0173924     1,25   0,210    -,0123061    ,0558709
dim_int_with_maxJacc |   -,001673   ,0166632    -0,10   0,920    -,0343323    ,0309863
                 d_R |   1,689695   ,3166075     5,34   0,000     1,069156    2,310234
              d_KIBS |   1,064061   ,2410161     4,41   0,000     ,5916777    1,536444
               _cons |  -2,369805   ,3218803    -7,36   0,000    -3,000678   -1,738931
------------------------------------------------------------------------------------
```

### b.3. Probit model with 'd_net3' as dependent variable

```
. probit d_net3          N_int_abvTHR jacc_max dim_int_with_maxJacc   d_R d_KIBS , robust

Iteration 0:   log pseudolikelihood = -107,23191
Iteration 1:   log pseudolikelihood = -76,040059
Iteration 2:   log pseudolikelihood = -73,799547
Iteration 3:   log pseudolikelihood = -73,770769
Iteration 4:   log pseudolikelihood = -73,770745
Iteration 5:   log pseudolikelihood = -73,770745

Probit regression                                 Number of obs   =        352
                                                  Wald chi2(5)    =      48,95
                                                  Prob > chi2     =     0,0000
Log pseudolikelihood = -73,770745                 Pseudo R2       =     0,3120

--------------------------------------------------------------------------------------
                     |               Robust
              d_net3 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------+----------------------------------------------------------------
        N_int_abvTHR |   ,0154214   ,0041721     3,70   0,000     ,0072442    ,0235985
            jacc_max |   ,0411871    ,018893     2,18   0,029     ,0041576    ,0782166
dim_int_with_maxJacc |  -,0203452    ,018721    -1,09   0,277    -,0570376    ,0163472
                 d_R |   1,914026   ,3642839     5,25   0,000     1,200043    2,628009
              d_KIBS |    ,953101   ,2931808     3,25   0,001     ,3784771    1,527725
               _cons |  -2,717697    ,370329    -7,34   0,000    -3,443528   -1,991865
--------------------------------------------------------------------------------------
```

1001010001100011110110101101010100111111010001101011100101001111111001001110001111010101010111111001

# References


- Amendola M. and Gaffard J. L. (1988), *The innovative choice,* Basil Blackwell, Oxford
- Amendola M., Gaffard J. L. (1998), *Out of Equilibrium*, Oxford, Clarendon Press
- Audretsch, D. (2002) *Entrepreneurship: A survey of the literature,* Bruxelles, European Commission, Enterprise Directorate General
- Bonifati G., Villani M. (2013), *Exaptation and emerging degeneracy in innovation processes,* Economics Of Innovation And New Technology, Vol. 22, No. 1
- Breschi, S., Malerba, F. (1997), *Sectoral systems of innovation: technological regimes, Schumpeterian dynamics and spatial boundaries,* in: Edquist, C. (Ed.), Systems of Innovation. Frances Pinter, London
- Caloffi A., Rossi F., Russo M. (2011), *Promoting successful innovation networks: a methodological contribution to regional policy evaluation and design. The case of Tuscany's innovation policies 2000-2006*, Dipartimento di Economia Politica, Università di Modena, WP 657
- Caloffi, A., F. Rossi, F., A., M. Russo (2013), *Networked by design: can policy constraints support the development of capabilities for collaborative innovation?,* Birkbeck College, University of London, Department of Management WP 5
- Caloffi A., Rossi F., Russo M. (2014), *What Makes SMEs more Likely to Collaborate? Analysing the Role of Regional Innovation Policy*, European Planning Studies
- Caloffi A., Rossi F., Russo M. (2015), *The roles of different intermediaries in innovation networks: A network-based approach to the design of innovation policies,* in Geyer, R. and Carney, P. (eds.) Handbook on Complexity and Public Policy
- Chen M., Kuzmin K., Szymanski B. K. (2014), *Extension of Modularity Density for Overlapping Community Structure*, in Proc. 2014 IEEE/ACM Int. Conf. Advances in Social Networks Analysis and Mining, Beijing, China
- Clauset A., Newman M. E. J., Moore C. (2004), *Finding community structure in very large networks*, Physical Review E, vol. 70, Issue 6
- Cover T., Thomas J. (2006), *Elements of Information Theory*, Wiley, Second Edition
- Esquivel A. V., Rosvall M. (2011), *Compression of flow can reveal overlapping modular organization in networks*, Physical Review X, vol. 1, Issue 2
- European Commission (2003), *Investing in research: an action plan for Europe*, COM(2003)226, Bruxelles: Commission of the European Communities
- European Council (2000), *Council Decision on a Multiannual Programme for Enterprise and Entrepreneurship, and in Particular for Small and Medium Enterprises*
- Filisetti A., Graudenzi, A., Serra R., Villani M., Füchslin R. M., Packard N., Kauffman S., Poli I. (2011), *A stochastic model of autocatalytic reaction networks. Theory in biosciences*, June 2012, Volume 131, Issue 2
- Filisetti A., Villani A., Roli A., Fiorucci M., Serra R. (2015), *Exploring the organisation of complex systems through the dynamical interactions among their relevant subsets*, Proceedings of the European Conference on Artificial Life 2015
- Fortunato S. (2010), *Community detection in graphs*, Physics Reports, vol. 486
- Freeman C. (1988), *Japan: A New National System of Innovation?*, in G. Dosi, C. Freeman, R. Nelson, G. Silverberg, L. Soete (eds.), Technical Change and Economic Theory. London and New York: Pinter Publisher

- Freeman C. (1991), *Network of Innovators: A Synthesis of Research Issues*. Research Policy, Vol. 20, Issue 5
- Georgescu-Roegen N. (1971), *The entropy law and the economic process*, Cambridge (USA): Harvard University Press, 1971
- Haberman S. J., (1973), *The analysis of residuals in cross-classified tables,* Bio- metrics, 29
- Hicks, J. R. (1970), *A Neo-Austrian Growth Theory*, Economic Journal, Royal Economic Society, vol. 80(318), June.
- Jaccard, Paul (1901), *Étude comparative de la distribution florale dans une portion des Alpes et des Jura*, Bulletin de la Société Vaudoise des Sciences Naturelles 37 Lambiotte R., Rosvall M. (2012), *Ranking and clustering of nodes in networks with smart teleportation*, Physical Review E 85
- Lane D. (2002), *Complexity and Local Interactions: Towards a Theory of Industrial Districts'*, in A. Quadrio Curzio and M. Fortis (eds.) Complexity and Industrial Clusters: Dynamics and Models in Theory and Practice. Heidelberg and New York: Physica-Verlag.
- Lane D. (2004), *Innovation led policies for clusters and business networks: Globalization, ontological uncertainty and degeneracy*, convegno OECD "Verso un'economia della conoscenza", Firenze
- Lane D. (2011), *Complexity and Innovation Dynamics*, Chapter 2 in Handbook on the Economic Complexity of Technological Change
- Lane D., Maxfield R. (1997), *Complexity, foresight and strategy*, in Arthur W., Durlauf S., Lane D., The Economy as a Complex Evolving System II, Redwood City, CA, Addison-Wesley
- Lane D., Maxfield R.(2005), *Ontological Uncertainty and Innovation*, Journal of Evolutionary Economics 15
- Lane D., Maxfield R. (2009), *Building a New Market System: Effective Action, Redirection and Generative Relationships*, in Lane D., West G., Van der Leeuw S., Pumain D., eds., Complexity perspectives on Innovation and Social Change
- Lane D., Serra R., Villani M., Ansaloni L. (2005), *A theory based dynamical model of innovation processes*, ComplexUs 2 (3-4)
- Lucas R. E. (1988), *On the mechanics of Economic Development*, Journal of Monetary Economics 22
- Lundvall B-Å. (1985), *Product Innovation and User-Producer Interaction*, Aalborg: Aalborg Universitetsforlag
- Lundvall B.-Å. (1988), *Innovation as an Interactive Process: from User-Producer Interaction to the National System of Innovation*, in Dosi G. *et al.* (eds.), Technical Change and Economic Theory (London: Pinter)
- Lundvall, B.-Å. (1992), *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*, London: Pinter
- Malerba F., (2002), *Sectoral systems of innovation and production*, Research Policy 31
- Malerba F., Orsenigo L., (1997), *Technological regimes and sectoral patterns of innovative activities,* Industrial and Corporate Change 6
- Mowery D., Teece D. (1996), *Strategic alliances and industrial research,* in R. Rosenbloom, W. Spencer (Eds.), Engines of innovation: U.S. industrial research at the end of an era. Boston, MA: Harvard Business School Press
- Nelson R. (1988), *National Innovation Systems: a Comparative Analysis*, Oxford University Press
- Nepusz T., Petróczi A., Négyessy L., Bazsó F., *Fuzzy communities and the concept of bridgeness in complex networks*, Phys. Rev. E, vol. 77

- Newman M. E. J. (2004), *Fast algorithm for detecting community structure in networks*, Physical Review E, vol. 69
- Newman M. E. J. (2006), *Modularity and community structure in networks*, Proceedings of the National Academy of Sciences, vol. 103, no. 23
- Newman M. E. J., Girvan M. (2004), *Finding and evaluating community structure in networks*, Physical Review E 69
- Palla G., Derényi I., Farkas I., Vicsek T. (2005), *Uncovering the overlapping community structure of complex networks in nature and society*, Nature 435
- Papoulis A. (1991), *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, New York
- Rebelo S. (1991), *Long-Run Policy Analysis and Long-Run Growth*, Journal of Political Economy 99 (3)
- Romer P. M. (1986), *Increasing Returns and Long-Run Growth*, The Journal of Political Economy, Vol. 94, No. 5.
- Rossi F. (2007), *Innovation policy in the European Union: Instruments and objectives*, MPRA Paper, Munich, Germany, University Library of Munich
- Rossi F., Caloffi A. Russo M. (2014), *Can policy design help organizations improve their networking capabilities? An empirical analysis on a regional policy*, Fteval Journal for Science and Technology Policy Evaluation
- Rossi F., M. Russo, S. Sardo, J. Whitford (2010), *Innovation, Generative Relationships and Scaffolding Structures*, in: P. Ahrweiler (Ed) Innovation in Complex Social Systems, London, Routledge
- Rosvall M. Bergstrom C. T. (2008), *Maps of information flow reveal community structure in complex networks*, PNAS
- Rosvall M., Esquivel A. V., Lancichinetti A., West .D., Lambiotte R. (2014), *Memory in network flows and its effects on spreading dynamics and community detection*, Nature Communications
- Russo M. (2000), *Complementary innovations and generative relationships: An ethnographic study*, Economics of Innovation and New Technology, 9
- Russo M. (2005), *Innovation processes and competition: China challenges Sassuolo. How do innovation processes in industrial districts are affected by competition from new market systems.* In Proceedings of the conference in honor of Professor Sebastiano Brusco, Clusters, Industrial Districts and Firms: The Challenge of Globalization, September 12–13
- Russo M., Rossi F. (2005), *Stimolare l'innovazione con strumenti innovativi: reti di partenariato e sviluppo locale nel PRAI-ITT 2002-03*, Rassegna Italiana di Valutazione, n. 32
- Russo M., Rossi F. (2007), *Politiche per l'innovazione: dalla valutazione alla progettazione*, Materiali discussione DEP, 565, Unimore
- Russo M., F. Rossi (2009a), *Innovation Policy: Levels and Levers*, in D. Lane, Pumain D., Van der Leeuw S., West G. (eds.) Complexity Perspectives in Innovation and Social Change, Berlin: Springer
- Russo M., F. Rossi (2009b), Cooperation Networks and Innovation: A Complex Systems Perspective to the Analysis and Evaluation of a Regional Innovation Policy Programme, Evaluation, 15 (1)
- Russo M., Whitford J. (2009), *Industrial districts in a globalizing world: A model to change, or a model of change*, Materiali discussione DEP, 615, Unimore
- Sardo S. (2009), *Brokeraggio tecnologico nel settore metalmeccanico in Emilia- Romagna: dal Parco Scientifico Tecnologico ex-SIPE a CRIT srl*, Materiali discussione DEP, 614, Unimore

- Saxenian A., (1994), *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*, Cambridge, MA: Harvard University Press
- Serra R., Villani M., Lane D. (2009), *Modeling Innovation*, Chapter 13, in Lane D., Pumain D., S. Van der Leeuw (eds): Complexity perspectives on innovation and social change. Berlin: Springer
- Shannon C.E. (1948), *A Mathematical theory of communication*, Bell Labs Tech J 27
- Shannon C.E., Weaver W. (1949), *The Mathematical Theory of Communication*, Univ. Illinois Press,Chicago.
- Shen H., Cheng X., Cai K., Hu M.-B., (2009), *Detect overlapping and hierarchical community structure in networks*, Physica A: Statistical Mechanics and its Applications, vol. 388, no. 8,
- Solow R. M. (1956), *A Contribution to the Theory of Economic Growth*, The Quarterly Journal of Economics, Vol. 70, No. 1
- Solow R. M. (1957), *Technical Change and the Aggregate Production Function*, The Review of Economics and Statistics, Vol. 39, No. 3
- Swan Trevor W. (November 1956), *Economic growth and capital accumulation*, Economic Record (Wiley) 32
- Tononi G., McIntosh A. R., Russel D. P. Edelman G. M. (1997), *Functional Clustering: Identyfing Strongly Interactive Brain Regions in Neuroimaging Data*, NeuroImage, vol. 7
- Tononi G., Sporns O., Edelman G. M. (1994), *A measure for brain complexity: Relating functional segregation and integration in the nervous system*, Proc. Natl. Acad. Sci. USA, vol. 91
- Tononi G., Sporns O., Edelman G. M. (1996), *A complexity measure for selective matching of signals by the brain*, Proc. Natl. Acad. Sci. USA, vol. 93
- Tononi G., Sporns O., Edelman G. M. (1998), *Measures of degeneracy and redundancy in biological networks*, Proc. Natl. Acad. Sci. USA, vol. 96, pp 3257-3262
- Villani M., Benedettini S., Filisetti A., Roli A., Lane D., Serra R. (2013a), *The detection of intermediate-level emergent structures and patterns*, in Liò P., Miglino O., Nicosia G., Nolfi S., Pavone M. (eds): Advances in Artificial Life, ECAL 2013. MIT Press. ECAL 2013 Best paper Award
- Villani M., Benedettini S., Roli A., Lane D., Poli I., Serra R. (2013b), *Identifying emergent dynamical structures in network models*, in Bassis S., Esposito A. and Morabito C. (eds): Recent Advances of Neural Networks Models and Applications. Proceedings of Wirn 2013. Springer series on Smart Innovation, Systems and Technologies 26
- Villani M., Filisetti A., Fiorucci M., Roli A., Serra R. (2015), *Exploring the organisation of complex systems through the dynamical interactions among their relevant subsets,* Proceedings of the European Conference on Artificial Life 2015
- Villani, M., Bonacini, S., Ferrari, D., Serra, R. (2009), *Exaptive processes: an agent-based model,* Chapter 15, in Lane D., Pumain D., Van der Leeuw S. (eds): Complexity perspectives on innovation and social change. Berlin: Springer
- Wakita K, Tsurumi T. (2007), *Finding community structure in mega-scale social networks*, Proceedings of the 16th international conference on World Wide Web, WWW '07
- Zhang S., Wang R.-S., Zhang X.-S. (2007), *Identification of overlapping community structure in complex networks using fuzzy c-means clustering*, Physica A: Statistical Mechanics and its Applications, vol. 374, no. 1