doi: 10.26398/IJAS.0030-005

A METHODOLOGICAL APPROACH TO INVESTIGATE INTERACTIVE DYNAMICS IN INNOVATIVE SOCIO-ECONOMIC COMPLEX SYSTEMS

Riccardo Righi[1]

European Commission, Joint Research Centre (JRC), Unit B6 - Digital Economy & CAPP – University of Modena and Reggio Emilia

Abstract. *Three aspects have been addressed to characterize agents' innovative interactions (Lane, 2011): relational structures, shared processes and common functions. A new methodological approach that allows their investigation is here developed. Each of the aforementioned aspects is separately analyzed through the implementation of one of the following community detection methodologies. Respectively, these algorithms are: Clique Percolation Method (CPM), Infomap (IM), and Relevance Index (RI). Areas of co-existence of the three aspects are investigated by considering those intersections determined by groups of agents that are simultaneously detected together by each of the three methodologies. Finally, the implementation of a linear regression model is used to analyze which types of interactions are associated with larger size of these intersections. This final analysis demonstrates that the proposed methodology leads to the identification of areas of the system in which statistically significant elements emerge. The approach is implemented in a case study regarding a cycle of policy interventions, developed in Region Tuscany (Italy) from 2000 to 2006, aimed at supporting innovative network projects among local economic agents.*

Keywords*: Innovation, Interactions, Community detection, Dynamic complex system*

1. INTRODUCTION

The present work aims to develop a statistical methodological approach that allows the investigation of the dynamics of socio-economic complex systems, in which innovative activities are developed, through the investigation of three separate aspects. The considered theoretical framework, which is based on the *ontological uncertainty* of innovation and which is focused on the investigation of the emergent features that typically lead to its occurrence, addresses *structures* and *processes* and

[1] Corresponding author: Riccardo Righi, email: riccardo.righi@ec.europa.eu; riccardo.righi@unimore.it. The views expressed are purely those of the author and may not in any circumstances be regarded as stating an official position of the European Commission.

functions as the most important aspects that characterize innovative dynamics (Lane, 2011; Lane and Maxfield, 1997, 2005). Since in this approach 'interactions' among agents represent the crucial dynamics that enable exaptation and disruptive changes, all the mentioned aspects concern 'agents developing relationships with other agents'. Thence, in order to take into account the presence of groups of agents in the investigation of areas of the system in which *structures* and *processes* and *functions* are present, community detection analyses are implemented. Firstly, specific algorithms are used to separately investigate the considered aspects, as represented in Figure 1. Then, the co-existence of *structures* and *processes* and *functions,* is analyzed by considering overlaps of communities detected by different algorithms, as represented in Figure 2. Qualitative patterns that emerge in these intersections, independently on the settings that are used for different implementations of the algorithms, demonstrate that the proposed methodological approach allows the grasp of evidences that are statistically far from randomness[2]. The original contribution of this work lies in the development of a combination of community detection analyses aimed at investigating agents involved in innovative dynamics. Additionally, an initial methodological step, based on the use of regression analysis, is made to describe the detected areas in which structures and processes and functions co-exist.

The paper is structured as follows. Considerations regarding innovation theories in economics, and the innovation approach considered for this work, are presented in Section 2. In Section 3, the methodological approach developed in the present work, is described. Then, in Section 4 a case study regarding a cycle of local public policies in sustain of network innovation projects is presented. The implementation of the three considered community detection analyses, each of which aimed at investigating one of the aforementioned aspects, is described in Section 5. Finally, in Section6, the analysis of the size of the resulting intersections among partitions (obtained with different methodologies) is investigated through the implementation of a linear regression analysis considering the types of interactions performed by involved agents.

[2] It is not an objective of this final part of the work to prove the validity of a specific regression model, or to find correspondence with elements addressed as relevant in literature. Rather, the final analysis regarding the size of the considered intersections is just intended as a means to demonstrate that the whole methodology allows the identification of areas of the system in which statistically significant elements are present.

2. A THEORETICAL APPROACH ABOUT INNOVATION

As long as innovation is considered as an exogenous determinant of economic growth, it is difficult to correctly understand the processes that facilitate it. Although mainstream economic theory has depicted innovation as an unforeseen shock, or as the results of a set of given resources/technologies, new approaches to the comprehension of innovative processes have identified specific conditions that typically lead to its occurrence. With the goal of deepening the comprehension of a phenomenon whose architecture was beginning to be seen as structured and complex, at the beginning of the 90's new contributions (Dosi et al., 1988; Lundvall et al., 1988; Nelson, 1993) started to investigate the role of the circulation of information and of interactions among involved agents (i.e. people, enterprises and local 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 meso-level of analysis (Breschi and Malerba, 1997; Etzkowitz and Leydesdorff, 2000; Malerba and Orsenigo, 1997; Saxenian, 1994) and at a micro-level of analysis (Freeman, 1991; Mowery and Teece, 1996; Russo, 2000). Moreover, the subject of innovation was also considered in the study of economic productive processes, and an *out of equilibrium* approach (Amendola and Gaffard, 1988, 1998) was able to create a gap with mainstream economics[3].

Instead of focusing on the prediction of the objects of innovation, academic personalities affiliated to the Santa Fe Institute (US) started to investigate the ontological aspects of innovation and to deepen the comprehension of the processes leading to its flourishing. In particular, the conceptualization of how innovation leads to a continuous re-definition of the use of present resources, therefore causing a continuous re-definition of the resources themselves, was developed. The inspiring contribution provided by David Lane, alone or with Robert Maxfield, (Lane, 2011; Lane and Maxfield, 1997, 2005), helped in the conceptualization of the *agent-artefact* space. A theory capable to describe the most important entities and the most important phases characterizing the formation of innovation, was so outlined. Far from looking for predictability, Lane's theory focuses on how the

---

[3] While in exogenous economic models and in endogenous growth models innovation is intended, respectively, as *"something that falls from the sky"* and as being constrained in an environment in which technologies and resources are given (Amendola and Gaffard, 1998), in the approach of Amendola and Gaffard innovation is intended as creating new resources (instead of being the result of them).

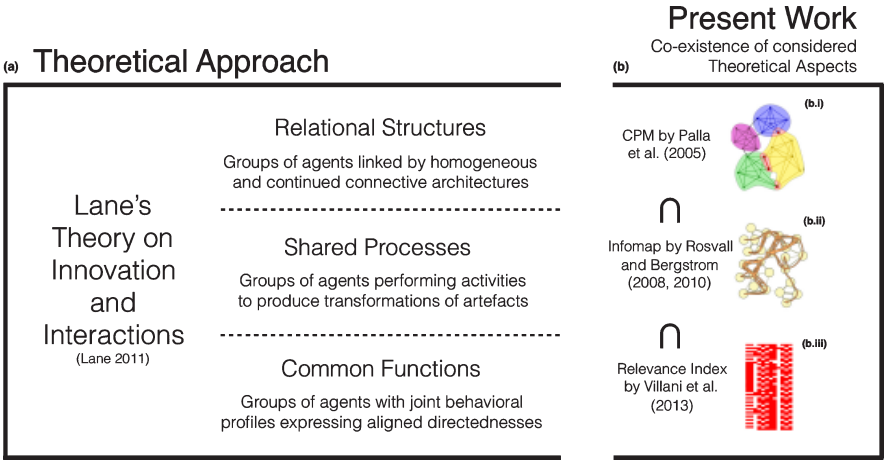


Fig. 1: (a) Lane's conceptualization regarding innovation processes. (b) Scheme of the present work: investigation of intersections (∩) generated by the groups agesents detected together by each of the three community detection methodologies. (b.i) Image representing CPM. Source: Palla et al. (2005). (b.ii) Image representing IM method. Source: Rosvall and Bergstrom (2008). (b.iii) Image representing RI method. Source: Villani et al. (2013).

processes that led to innovations are intertwined, making some recurrent elements emerge. Among those having a crucial impact on the likelihood to generate innovations in geographically defined systems, the presence of a high level of interactions among agents is emphasized, and the presence of three aspects is highlighted. These elements are: *structures*, intended as connections and relationships through which agents can interact; *processes,* intended as activities in which agents co-participate and through which they perform transformations of artifacts; and *functions,* intended as the ultimate and final goal towards which agents tend, therefore providing directedness to their joint actions.

Since the subject to be investigated deals with phenomena that do not concern 'agents' in their individual dimension, but concern 'agents' as part of interacting groups, community detection analyses are implemented[4] . The representation of the conceptual framework here adopted, is presented in Figure 1: a specific com munity detection algorithm is associated to each of the three aspects, and their coexistence is finally investigated by considering areas of entire intersection among communities obtained with the three methodologies. Only the investigation of the co-existence

[4] The presence of relational structures and the co-participation in processes and the pursuing of a same function, all concern a multiplicity of agents. The considered aspects cannot be referred to agents in their individual dimension.

of the aforementioned aspects is the final objective of this work, as it is addressed by the literature as the most relevant condition that deserves to be investigated.

3. THE METHODOLOGICAL APPROACH

In this section, the methodological approach is described. In Subsection 3.4, the selection of three community detection algorithms, as presented in Figure 1 and Figure 2(b), is discussed. The selection of these methodologies is developed as an evaluation regarding the theoretical coherence between the functioning of the algorithms and the specific objects to be investigated. Then, by taking into account one combination of obtained partitions at a time (one partition per each community detection methodology, see Figure 2(c)), the map of overlaps among communities detected with different methodologies is obtained[5], as presented in Figure 2(d). In Subsection 3.5, the focus is finally pointed into areas made of agents that are grouped together by all the methodologies (black areas in Figure 2(e)) as, in them, structures and processes and functions co-exist. The identification of these areas of entire intersection, is followed by the investigation of what concerning them. Two statistics are developed in order to further analyze the size of the resulting intersections. Before discussing the selection of the community detection algorithms, three considerations are discussed in the following.

3.1 COMMUNITIES AS EMERGING ENTITIES

The aim of this work is to investigate multiple aspects of agents' interactions, in a real socio-economic complex network in which affiliations to actual communities are not present. Communities are here intended as emerging entities that have to be unveiled. And the statistical methodologies are implemented in order to make them emerge and to capture specific aspects (see Section 2) in socio-economic dynamic complex system. Therefore, it is not an objective of this work to compare the efficacy of different community detection algorithms.

---

[5] The four kinds of areas of overlap are represented with different colors: orange for the overlaps between a community of CPM and one of IM, green for the overlaps between a community of CPM and one of RI, purple for the overlaps between one community of IM and one of RI, and finally black for the complete intersections among one community of CPM and one of IM and one of RI. Apart from the colors, different overlaps can be identified by considering the types of the communities involved, as indicated in the labels. Figure 2(d) has to be interpreted as a series of Venn diagrams in which the crucial element is how the methodologies allow the identification of qualitatively different areas of overlap.

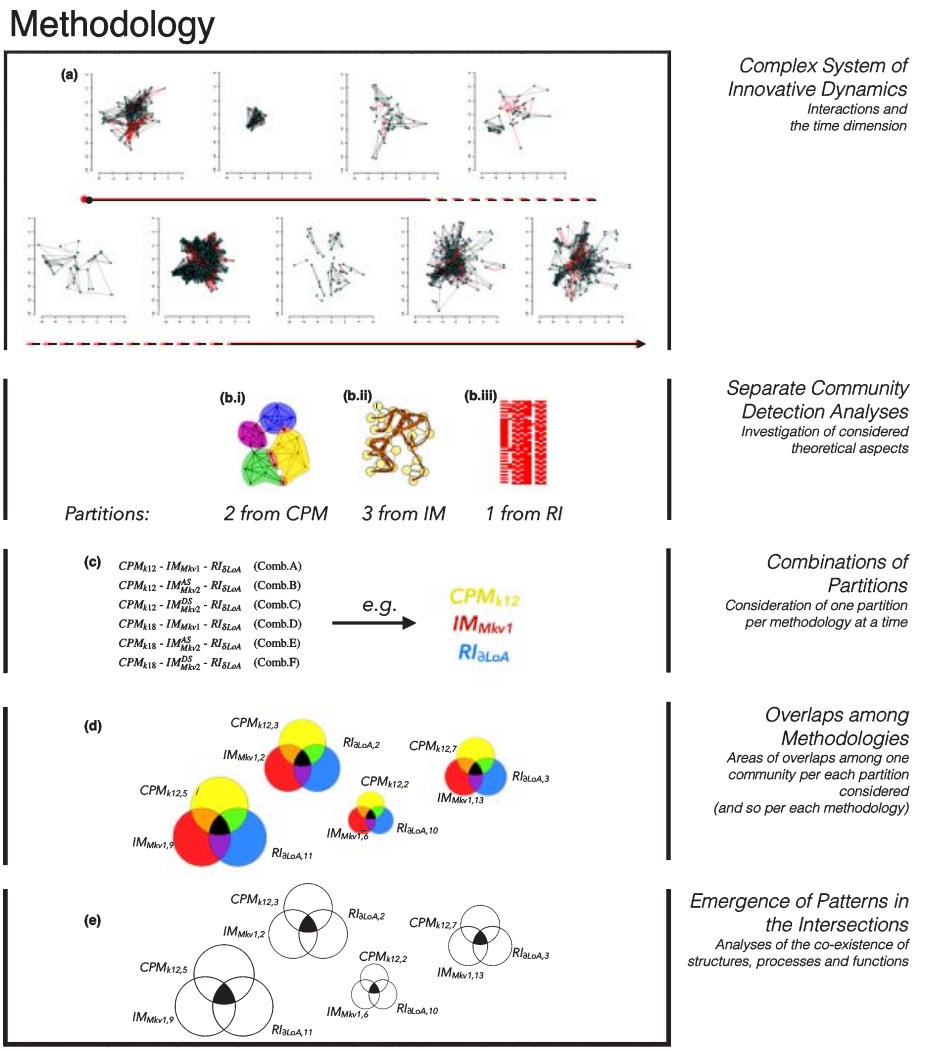


Fig. 2: Scheme presenting the methodological implementation proposed in this work. In (a), representation of a dynamic complex system. Nodes indicate agents, edges represent interactions. The red horizontal arrow represents time. Each node maintains always the same position in all the graph and is represented only if active. Fruchterman Reingold layout. Source: personal elaboration on Region Tuscany Network Policies 2000-2006, in which the snapshots represent the nine waves of the whole policy cycle. In (b.i)-(b.iii), the three different community detection methodologies used, as presented in Figure 1. In (c), list of the combinations of partitions obtained from the different settings of the different community detection analyses, with reference to the case study considered in this work (see Subsection 6.2). In (d), exemplification of overlaps among communities that are included in partitions obtained with different community detection algorithms. In (e), representation of the intersections, i.e. the areas of co-existence of *structures, processes* and *functions.*

3.2 TIME DYNAMICS

The investigation of interactive and innovative processes requires an adequate consideration of the time dimension. The adopted theoretical approach about innovation considers, as unit of analysis, agents that reiterate interactions over time. Some considerations follow this. First of all, the network representing the considered system cannot be intended just as a series of instants which can be independently analyzed; rather, as a system driven by a whole and unitary evolution. Even if instants can be captured, their analyses cannot be detached to those regarding the other instants[6]. The approach here developed acknowledges the timecontinuous dimension of the dynamics of reproduction[7]. Therefore, in order to investigate the connection among the past and the present and the future, the observed phenomena cannot be easily subsetted: even if some relevant sub-periods can be identified, the weave of these dynamics must not be broken. Finally, exclusively in the investigation of the aspect regarding the structures, time dynamics will not be observed. In this analysis, the entire resulting network will be considered, without taking into account information on its evolution[8]. Here the time will be intended as the final sum of what occurred.

---

6 In the considered case study, even if the interactions can be subsetted by using the policy waves as units of time, the whole cycle has to be intended as unitary. The subsetting of time can be properly implemented only if a deep comprehension of what then considered is present. In the case study of Tuscany, the one taken into account in this work, the policies that sustained the waves were dealing with a problem of allocation of resources over a substantial period of time. Some priorities, for instance, were established in different periods depending on the ongoing solicitations coming from the economic, social and political environment. Moreover, the attempt to separately analyze sub-periods can immediately end because of a high degree of discontinuity in the resulting sequence. For instance, as presented in Figure 2(a) regarding the case study below considered, because of the consideration of sub-periods, the system appears as far from being driven by stepwise changes. Time is here represented as the sequence of what occurred in the waves, that from the point of view of the policy are the most crucial elements that can be considered. However, it can be observed how the obtained networks dramatically change from one snapshot to the other, both in terms of size (number of agents involved) and of structure (configuration of the partnerhips).

7 In economics, the dimension of reproduction was considered and investigated since the research field was born. The physiocratics (Quesnay, 1758) and the classics (Marx, 1867; Picchio, 1992; Sen, 1985, 1987; Smith, 1761, 1776) consider the economic systems very similarly as biologists and physicists. In these fields, the subject of the research are entities that have to preserve, survive and reproduce themselves over time. Among what distinguishes economics from other fields, there is the presence of normative, ethical and political aspects that the agents can engage, in order to orientate the evolution of the system they populate.

8 In this part of the analysis, the focus is on the final set of all the relationships that occurred, therefore considering the permanent presence of any relationship after its occurrence. In this way, interactions are intended as creating structures that persist after their activation

3.3 ALGORITHMS TO DETECT OVERLAPPING COMMUNITIES

The selection of community detection methodologies is restricted to those allowing the detection of overlapping communities. Since the role of bridging agents (also defined as 'brokers') in socio-economic systems has been widely recognized (Burt, 1995), any community detection analysis that necessarily produce exclusive affiliations (i.e. overlaps among communities not allowed) is not taken into consideration. However, even if overlaps are allowed, it is not the aim of this work to analyze the presence of overlaps among communities detected by the same methodology. Therefore, overlaps among communities detected with the same methodology are admitted but not analyzed. In the last part of this work (see Subsection 3.5), intersections are computed among combinations of three communities, each of which resulting from the implementation of a different methodology. In this way, the co-existence of the three aforementioned aspects is investigated.

3.4 A SPECIFIC COMMUNITY DETECTION ALGORITHM FOR EACH CONSIDERED ASPECT

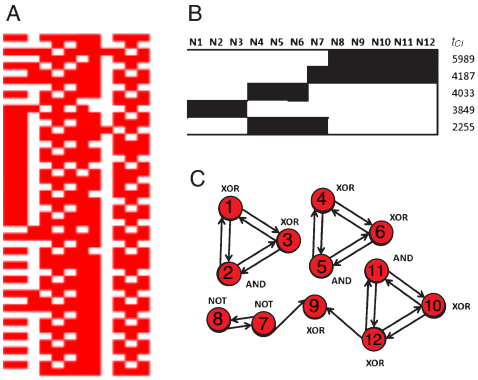
After having described these preliminary elements, the discussion regarding the coherence in the selection of each methodology is presented. This step is developed as an evaluation between the specific subject to be investigated and the functioning of the considered community detection algorithm. In order to coherently investigate the three aspects of innovative groups of agents, the implementation of the three following methodologies is proposed:

- the Clique Percolation Method (henceforth, CPM) (Palla et al., 2005) to investigate how agents established specific architectures of relationships (*relational structures*);
- Infomap algorithm (henceforth, IM) (Rosvall and Bergstrom, 2008, 2010; Rosvall et al., 2014) to investigate exchanges of information as determinant flows of co-participated economic initiatives (*shared processes*);
- the Relevance Index algorithm (henceforth, RI) (Filisetti et al., 2015; Roli et al., 2017; Villani et al., 2013) to investigate the entropy of agents' behaviors over time, as an element able to unveil agents' similarity in terms of their role in the system, and in terms of the goals towards which their action is directed (*common functions*). In Figure 3, an example of how the algorithm works is proposed.

The first methodology, CPM, is based on the detection of adjacent k-cliques, where a k-clique is a complete subgraph made of k nodes, and where to be 'adjacent' means to share k–1 nodes (Palla et al., 2005). Each node of a generic detected community $CPM_{k, a}$, where k is the size of the fundamental clique and a indicates

one of the communities in the resulting partition, is connected with at least k–1 nodes of the same community. As the clique 'rolls' over the network, each detected community presents a continuously connected architecture. Even in the case of a community having low internal edge density, a 'chain' of cliques is necessarily present (Fortunato, 2010), hence providing continuity to the architecture of the community. Therefore, since the subject to be investigated is the aspect regarding the presence of relational structures within groups of agents, the CPM is suggested as a coherent and appropriate methodology.

The second methodology, IM, is based on the minimization of the two-level binary description of a simulated flow over the network (Huffman, 1952; Rosvall and Bergstrom, 2008, 2010; Rosvall et al., 2014). The community detection is performed through the identification of groups of agents, within which the simulated flow tends to circulate for a certain period before exiting. Since the context of analysis is a socio-economic environment, this simulated flow is addressed as a flow of information moving among the agents of the system. Assuming this interpretation, the detection of areas in which the flow tends to stay longer before exiting, indicates groups of agents intensively sharing information. Based on the robust economic literature regarding the role of information and the functioning of processes of production (Amendola and Gaffard, 1988, 1998; Coase, 1937; Georgescu-Roegen, 1971; Hicks, 1973), the presence of joint information has to be considered as a crucial element in the development of an economic activity. Since the development of an economic process and the spread of concentrated flows of information are connected, IM can be considered as a meaningful methodology to detect groups of agents sharing economic processes.



(Villani et al. 2013)

Fig. 3: Figures presenting the functioning of RI algorithm. In (A), matrix describing the boolean states (red equals to presence of activity, white inactivity) of a set of 12 agents over 48 instants in time. Each row corresponds to an instant, and each column to an agent. In (B), the t_{CI} of the best groups detected. Black cells indicate the agents (columns) included in the detected groups (rows). In (C), the boolean network that has been used to gen erate the simulated behavioral profiles of activity shown in (A). Functional groups are detected through the consideration of entropy-based measures in joint dynamic behaviors. Source: (Villani et al., 2013).

Finally, RI methodology is a algorithm that considers how agents' behaviors show levels of integration over a period of time, and allows the identification of functional subsystems. In the inspiring contribute of Tononi (Tononi et al., 1994), information-theory measures (entropy, integration and mutual information) were used to investigate the functioning of neurons in brain regions, independently from their anatomical proximity[9]; the presence of integration among specific neurons' activities' confirmed the hypothesis that these neurons had similar functions (Tononi et al., 1998). The same statistical approach, based on entropy measures, has been implemented in the algorithm initially called Dynamic Cluster Index (DCI) and then ameliorated in a following version called RI (Roli et al., 2017; Sani et al., 2016; Villani et al., 2013). The algorithm is schematically presented in Figure 3. RI is adapted for implementation in socio-economic complex systems[10], in order to investigate the presence of communities whose agents are characterized by similar objectives and purposes (Righi et al., 2017).

3.5 DETECTION OF THE CO-EXISTENCE OF *STRUCTURES, PROCESSES* AND *FUNCTIONS*

The final part of this work investigates the presence of specific patterns, by considering the characteristics of the subsets of agents that are simultaneously grouped jointly through all the considered methodologies. Since community detection analyses are used to separately investigate the aspects of *structures* and *processes* and *functions,* the final co-existence of the three aspects can be assessed by considering entire intersections. Therefore areas determined by the overlap of a combination of communities in which (i) the first community is one of the communities detected with CPM, and (ii) the second one is one of the communities detected with IM, and (iii) the third one is one of the communities detected with RI, are considered.

Then, the size of the obtained intersections is analyzed. The objective is to comprehend if there are elements that are correlated with the number of agents that

---

[9] Tononi assumed that neurons with similar functions have high level of coordination in their behaviors over time, independently from being or not located within the same brain region. In the field of neurological studies, two theories have always been opposed: the first, a localizationist theory sustaining that the brain is divided into separate areas characterized by specific functions; the second sustaining the presence of a holistic scheme of the 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.

[10] Even if the algorithm was implemented initially in the field of physics and biology, and tested particularly in random boolean networks, catalytic reaction models and in biological gene regulatory systems (Filisetti et al., 2015), recent works suggest to proceed in the evaluation of physical order and entropy, in order to assess how and why information is used in socio-economic complex systems (Hidalgo, 2015).

are included in these areas[11]. In this way, a crucial issue, namely the size of the areas in which the aforementioned aspects co-exist, is analyzed. Two statistics are computed. The first statistic, namely s_1 , is defined as the number of agents (i.e. the cardinality of the intersection) that are simultaneously grouped together, by each of the three aspects considered in the approach. This is computed with the following equation:

$$s_1 = \left| X_{CPM} \cap Y_{IM} \cap Z_{RI} \right| \tag{1}$$

where

- X_{CPM} is one of the community detected in the considered CPM partition;
- Y_{IM} is one of the community detected in the considered IM partition;
- Z_{RI} is one of the community detected in the considered RI partition.

Additionally to the consideration of the cardinality, the size of the intersections can be estimated also based on the sizes of the communities that generated it. An intersection obtained from three communities that are very large and different among them, cannot be considered as equal to an intersection with the same cardinality, but resulting from three communities each of which is not much larger than the intersection itself[12]. Therefore, a second statistic measuring how much an intersection corresponds to the three communities generating it, is considered. This statistic, namely s_2 , is computed as follows:

$$s_2 = \frac{\left| X_{CPM} \cap Y_{IM} \cap Z_{RI} \right|}{\left| X_{CPM} \cup Y_{IM} \cup Z_{RI} \right|} \tag{2}$$

where X_{CPM} and Y_{IM} and Z_{RI} are defined as in Equation 1. To evaluate the correspondence between the three communities and their intersection, the ratio between the cardinality of the intersection and the cardinality of the correspondent union set, is taken into account.

These two statistics, i.e. s_1 and s_2, are considered as dependent variables in linear regression models, which aim to assess if and how the size of the intersection is characterized by not random elements. The idea is to investigate what can be associated with the presence of larger overlaps among the results of the considered

---

[11] Since interactions are intended as dynamic phenomena that involve multiple agents, only the intersections that involve at least two agents are taken into consideration.

[12] Since here intersections are taken into account, only a part of each of the communities that generates them, is considered: each of the involved communities loses its unitarity. Therefore, a measure of correspondence between the intersections obtained and the communities generating them, is necessary.

methodologies. How the analyses can be implemented depends on the specific case study[13]. The models that are specifically developed for the case study considered in this work are presented in Section 6.

4. THE CASE STUDY: REGION TUSCANY POLICIES IN SUSTAIN OF INNOVATION

Analyses have been realized thanks to an original and unique dataset of network projects developed in the context of a cycle of public policies exclusively aimed at financing innovative proposals, and funded by regional government of Tuscany (Italy) from 2000 to 2006 (Caloffi et al., 2015). In order to participate and to receive fundings, agents had to develop projects in collaboration with other agents (the policies allowed exclusively the granting of fundings to partnerships of agents). The cycle of public policies was composed of nine waves 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. In Figure 2(a) the interactions occurred within each of the nine waves are presented in the one mode graphs of co-participations (the edges) of agents (the nodes), where the red horizontal arrow indicates their chronological order (with regard to the beginning of the first project). Grey edges represent co- participations in funded projects, while red edges represent co-participations in projects that were not fund. Since in each of the nine graphs each agent maintain always the same coordinates (and since agents are shown only if they participated in the corresponding wave), it possible to observe how the system had highly discontinuous changes over time (Righi et al., 2017), both in terms of number of agents involved and in terms of relationships activated.

The case study can be conceptualized as a bipartite dynamic network in which agents participated in funded and not funded projects[14]. Since the majority of agents participated in just one project within the whole policy cycle, the analysis is restricted to those agents that at least participated twice (two funded projects). In this way, 352 agents considered,

[13] The final regression analysis quantitatively investigate a qualitative element: the types of interactions that determine larger intersections. This is what is developed in Section 6. However, it is not an objective of this work to develop a specific model and to generalize its results. Rather, the objective of this work is to prove that the intersections resulting from the implementation of the whole methodologies present characteristics that are not random.

[14] CPM and IM have been run exclusively on the network of the funded projects. The information concerning non funded projects is used for the identification of more detailed behavioral profiles in the implementation of RI methodology.

and 298 projects are considered (168 of them were funded, while 130 were not funded). Each project is characterized by a starting date and an ending date. For non funded projects the considered starting date is the deadline for the presentation of the proposal of the project, and the considered ending date is the day after the deadline[15]. Thanks to this information, it is possible to extract 59 instants over time in which there are changes in the set of connections[16]. Finally, it is important to remark that, even if available, the information regarding the technological domain characterizing each project is not taken into account for the implementation of the community detection analyses. Nevertheless, this information is used in the last part of the work to implement regression models aimed at characterizing the kinds of activities developed by the agents belonging to the intersections generated by the overlaps of the three methodologies (see Section 6).

5. THE IMPLEMENTATION OF THE METHODOLOGY

All the methodologies are implemented over the same complex system, made up of the 352 agents that participated at least in two projects of the Region Tuscany policies in sustain of innovation. Depending on the kind of algorithm implemented, the time dimension is adequately treated.

5.1 CPM: AN A-POSTERIORI SELECTION OF THE VALUE OF *K*

The informational basis used for the implementation of CPM methodology is the binary network of all the connections that occurred among agents, independently on when they occurred. The bipartite network of agents and projects (see Section 4) is converted in the unweighted one-mode projection of the network of agents. In order to run the CPM algorithm, all the admissible values of *k,* regarding the size of the clique that 'rolls' through the network, are tested. The corresponding resulting partitions are evaluated a-posteriori, and just some of them are selected.

Figure 4 shows five statistics regarding the CPM partitions obtained with the admissible values of *k* (values of *k* represented on the x-axis and, on the y-axis, the scale of the values). Looking at these statistics, three ranges for the value of *k* are identified (in Figure 4, these ranges are determined by vertical red dashed lines). The first group of partitions is associated with values of *k* included between 3 and 8: all these partitions are characterized by the presence of one community having

---

[15] It was evaluated as important to consider also non funded projects, since interactions for the construction of each proposal were developed.

[16] Since each date concerns the starting or the ending of at least one project, on each of them at least one change in terms of ongoing partnerships occurred.

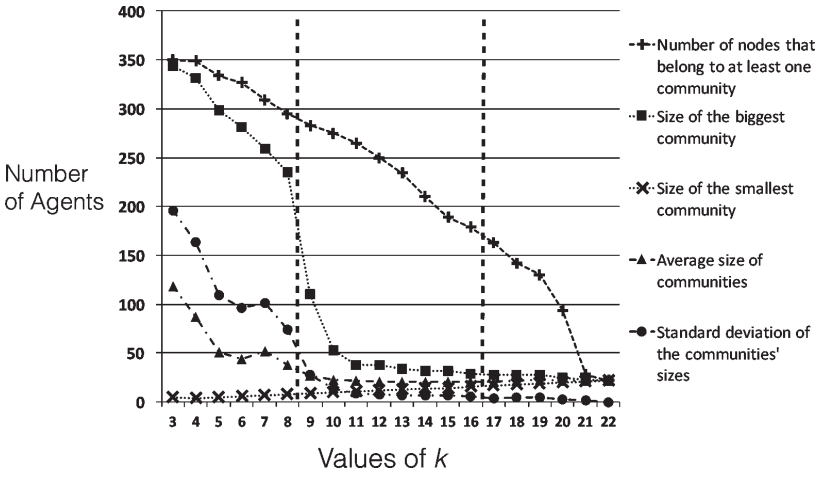


Fig. 4: Statistics describing the communities belonging to the CPM partitions obtained with different values of *k*.

cardinality larger than 225 (agents). Since the number of agents overall included in the considered network is 352, it is meaningless to consider partitions dominated by the presence of such a large and overwhelming community. Therefore, the values of k between 3 and 8 are not taken into account for the continuation of the analysis. The limit between the first and the second range is set by considering the fall in the size of the biggest community (in Figure 4, line with squared marks) between $k = 8$ and $k = 9$. The limit between the second and the third range is set by considering the large decreasing of the number of agents included in at least one community (in Figure 4, line with cross marks) when moving from $k = 20$ to $k = 21$. Since the number of agents included is meaningless for values of k higher than 20, the limit between the second and the third range is set in order to have, in the last range, partitions including nearly 150 agents.

After the definition of these three ranges, the first of which is completely excluded, only one representative value of k is selected in the remaining. For the range regarding values between $k = 9$ and $k = 16$, the partition taken as representative[17] is the one obtained with $k = 12$, henceforth named CPM_{k12}. For the range regarding values between $k = 17$ and $k = 22$, the partition taken as representative[18] is the one obtained with $k = 18$,

[17] This selection is made considering the fact that, after the fall in the size of the biggest community (line with squared marks), starting from $k = 11$ the same statistic flattens at a value of almost 40 agents. The intention is to capture the first partition in which no differences are present (in terms of the size of the biggest community) with regard to the preceding partition. Therefore, since still a variation between $k = 10$ and $k = 11$ is observed, $k = 12$ is considered.

henceforth named CPM_{k18}. Statistics of the considered CPM partitions are presented in columns 1-2 of Table 1.

5.2 INFOMAP: DIFFERENT HYPOTHESES TO SIMULATE FLOWS OF INFORMATION

The second methodology considered is IM (Rosvall and Bergstrom, 2008, 2010), which allows the detection of communities within which flows are fluent: exchanges between communities happen, but the majority of flows occur within each of them. The informational basis that is used is the weighted (by the number of co-participations in projects) one-mode projection of the network of the agents. Moreover, in order to consider the observed time dynamics (Rosvall et al., 2014), the chronological sequence of the projects is used to constrain the circulation of the simulated flow; and since different considerations can be made, different community detection analyses are developed. Therefore, different partitions are finally obtained. In each of the analyses, IM is set to detect overlapping communities, and the teleportation parameter equals to 0.15.

The first partition proposed (henceforth, named IMMkv1) is the result of an analysis in which the Markovian order is set to be equal to 1. This means that no information about the provenance of the flow is taken into account. When the flow moves through the network, in each step it faces a set of probabilities to move towards different directions that depend on: (i) the position of the flow; (ii) the weights of edges (given by the number of co-participations between agents). To simulate the propagation of a flow over the whole network without any kind of constraint is like observing the propagation of an information stream in a context where all finally observed edges are open. The flow runs in the network that is observed in the end of the policy cycle under the assumption that, after the collaborations in the specific project, the agents maintained all the relationships developed as active.

The second and the third analyses developed (henceforth named IM^{DS}_{Mkv2} and IM^{AS}_{Mkv2}) consider a second order Markov condition. The direction that the simulated flow takes at each step depends also on where it comes from. The process of creation of the simultated flow is implemented through the creation of *trigrams*[19] (Rosvall

---

[18] Here the statistic taken into account for the most, is the number of agents included in at least one community (line with cross marks). This statistic decreases almost constantly since the smallest *k*. Its largest fall is observed between $k = 20$ and $k = 21$, and a gap bigger than usual is observed also between $k = 19$ and $k = 20$. Therefore, in order to not select a k corresponding to a partition that is on the edge of a change, $k = 18$ has been chosen as the value representative of this range.

et al., 2014). In this context of analysis, the flow among ordered groups of three agents is supposed to be present in these two circumstances: (i) all three agents of the trigram are involved in the same project (*within project flows*); (ii) the first agent and the last agent participate in two different projects, while the second agent participates in both of these20 (*between projects flows*).

In order to model the circulation of information from an agent participating in one project, to an agent participating in another one (*between projects flows*), time can be taken into account. More specifically, the sequence in which projects occurred is the element that has to be considered. Since the flow can represent the information that is accumulated through the participation in projects, a first hypothesis is that the flow, through a single agent, can move from one project to another only if the first project ends before the beginning of the second. A second hypothesis is that, even if a project has not ended, agents involved in it can produce a flow of information moving outside the partnership of the project[21]. According to these considerations, two community detection analyses with a second order Markov condition are so implemented. The name of the first is IM^{DS}_{Mkv2}, that stands for: InfoMap with Markov order 2 and Denied Simultaneity[22]. The second, that is less restrictive than the previous, is IM^{AS}_{Mkv2}, that stands for: InfoMap with Markov order 2 and Admitted Simultaneity. Figure 5 presents an example of the consideration of different hypotheses on a second order Markov condition, as they have been taken into account in this work. The characteristics of the three partitions detected with IM algorithm, are presented in columns 3-5 of Table 1.

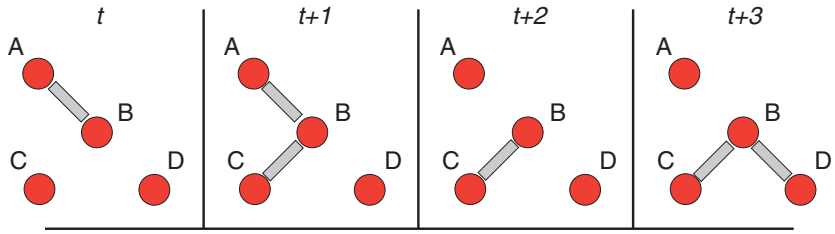
---

19 The trigrams are ordered sequences of three nodes describing two movements of a flow. If *A, B* and *C* are three agents (nodes) and the considered trigram is $A \rightarrow B \rightarrow C$, the movements of the flow that are considered are (i) from *A* to *B,* and then (ii) from *B* to *C.*

20 The second agent, which is the one that participates in both the two projects, acts as a 'bridge' between the first and the third agents. Therefore, he also acts as a bridge between the project in which he is participating with the first agent, and the project in which he is participating with the third agent.

21 Underlying the first hypothesis, i.e. the one admitting *between projects* -only after the end of the first project, there is the idea that the circulating information concerns the experiences and the knowledge accrued thanks to the complete realization of the participation in the project itself. Only finishing a participation in a project, a participant can accumulate knowledge, so becoming able to give rise to a flow of information. On the other hand, admitting the possibility of a spreading of information before a project has terminated, as it has been taken into account in the second hypothesis, deals more with a concept of acquaintanceship, rather than of accumulation of knowledge.

22 The term 'simultaneity' refers to the agents' capability to generate flows of information from projects that still have not ended. If the simultaneity is admitted, it means that the agent can generate a flow between two projects in which is participating in, even if they are still going on.



A,B,C,D are the agents of the system
Each grey edge represents an active co-participation in a project
t is the time

Fig. 5: Example of the computation of trigrams depending on the Markov condition and on the considerations developed in this work. In IM^{DS}_{Mkv2}, the trigrams would be: A →B →C; and A →B →D. In IM^{AS}_{Mkv2}, the trigrams would be: A →B →C; C →B →A; C →B →D; D →B →C; and A→B→D.

5.3 RI ANALYSIS AND AGENTS' INDIVIDUAL STATES OF ACTIVITY OVER TIME

Through the consideration of how agents' behaviors show levels of integration over a considered period, RI methodology allows the identification of functional communities. By making use of Shannon entropy (Shannon, 1948), and of the related measures of integration (henceforth, I) and of mutual information (henceforth, *MI*), Tononi et al. (Tononi et al., 1998, 1994) introduced the concept of Cluster Index (henceforth, *CI*), for measuring the integration of the activities of neurons in brain regions[23]. The level of significance of the *CI*, namely t_{CI} , is the value according to which the detection of groups of agents is performed. Thanks to the work of Villani et al. (Roli et al., 2017; Villani et al., 2013), an algorithm that, based on t_{CI} statistic, allows the detection of functional clusters in random boolean network, has been developed: the Relevance Index algorithm (*RI*). Since the higher the t_{CI} the farther the joint behaviors are from randomness, the methodology allows the detection of groups of agents whose actions show integration. Given this aligned directedness, it is possible to infer as the detected groups of agents, with regard to

[23] *CI* makes possible to measure how much a group of agents presents behaviors that are integrated among them and that are scarcely integrated with those of other agents. Formally *CI* is defined as $CI(X) = I(X)/MI(X, U \setminus X)$, where U is the whole considered dynamic system, and $X \subset U$. Since I and MI values depend on the size of the subsystem that is under analysis, a homogeneous system is used to normalize these values and to measure the statistical significance of the *CI* of the considered subset X. A homogeneous system is a system having the same number of agents of the system to which it is referred, and each agent has a random generated behavior in accordance with the probability of the states it assumes in the reference system.

the functioning of the whole system of which they are part, have similar functions.

The information regarding agents' levels of activity allows the definition of a series of variables describing the variation of these levels over time[24]. These variables are generated so that they assume four different values according to how the number of active participations varies in time[25]. Agents' activity is not described just for what is in each instant, but for what it is in the present, conditioned to what it was in its nearest past. Therefore, a second order Markov condition is taken into account. Since the process is developed considering variables which concern the variations of the Levels Of Activity, the corresponding partition will be henceforth named $RI_{\delta LoA}$.

In order to implement RI in a case study with many agents and few instants (in the present case study, there are 352 agents and 59 instants in time), some heuristics are required[26]. Since the first run of the algorithm over the entire system has provided the detection of a group of agents which had a very scarce number of activities[27], these agents are dropped, and the analysis is repeated without considering them. This procedure of skimming of the best detected group is reiterated, until a group of scarcely active agents is no longer detected as 'best' group. Then the resulting groups are analyzed through a cluster analysis for binary

---

[24] The most important aspect for the implementation of RI in the considered case study concerns the informational basis used to describe agents' statuses of activity. As the methodology is based on the computation of joint entropy measures over time, the only meaningful time instants to be considered are those in which at least one variation (in terms of the statuses of activity of the agents) occurs. Therefore, in order to have the largest number of meaningful instants in time, the 59 different dates of starting and ending of projects (see Section 4) are used. Thanks to them, it is possible to define a profile of activity for each of the agents involved in the six years of the policy program: in every instant is considered in how many projects each agent is participating. Participations in non funded projects are considered to be active for just a single day: the day in which the projects of the same wave started.

[25] The four situations described by the considered variables, are: (i) the agent is not participating in any project (no activity); (ii) the agent is participating in a number of projects that is higher than the number of projects in which it was participating in the previous instant (increasing activity); (iii) the agent is participating in a number of projects that is equal to the number of projects in which it was participating in the previous instant (constant activity); (iv) the agent is participating in a number of projects that is lower than the number of projects in which it was participating in the previous instant (decreasing activity).

[26] Up to now RI has been developed and tested in research areas of random boolean network models, of catalytic reaction systems and of biological gene regulatory systems (Filisetti et al., 2015; Villani et al., 2013). In these researches, the considered sets of agents could be observed for a relatively high number of instants in time.

[27] As these agents participated in few projects, they generate low levels of entropy. Therefore, they are immediately recognized as extremely integrated. However, this integration is caused by the fact that they did almost anything.

data, aimed at uncovering similarities among them. After 6 rounds of analysis over sets of agents progressively skimmed from the best groups of agents detected in the pre- ceding rounds, in the 7-th round the algorithm detected, as best group, a small group of agents whose profiles of activities were intense. Then, from the whole set of groups identified in the 7-th round, 16 clusters of groups are detected and, for each of them, the group showing the highest tCI is taken as the representative of the cluster in which it is included[28]. In the end, each of the groups representing the obtained clusters, is intended as a RI community. Therefore, in the partition $RI_{\delta LoA}$, 16 communities are detected. The descriptive statistics of this partition are presented in column 6 of Table 1.

Tab. 1: Descriptive statistics of the partitions obtained.

Partitions	CPM_{k12}	CPM_{k18}	IM_{Mkv1}	IM^{AS}_{Mkv2}	IM^{DS}_{Mkv2}	$RI_{\delta LoA}$
Number of agents belonging to at least 1 community	250	142	352	352	352	239
Number of agents belonging to more than 1 community	119	45	94	199	218	201
Number of communities	27	10	23	54	62	16
Size of the biggest community	38	28	71	61	50	90
Size of the smallest community	12	18	2	2	2	4
Average size of communities	17.62	20.20	23.43	14.13	13.02	56.06
Standard deviation of the size of communities	6.55	2.85	16.62	12.73	12.37	33.91

[28] Since the groups identified in a single round are 15.000 and can differ among them just for the presence/absence of a single agent, the heuristic has been developed as follows: (i) run of the algorithm and final ranking of all analyzed groups; (ii) selection of the groups with the highest tCI value, i.e. the 'best' group, in the end of every round of analysis; (iii) progressive skimming, round after round, of the best group that is detected at the end of the preceding round; (iv) stopping of the skimming procedure when the researcher evaluates that the best group of agents should not be excluded; (v) hierarchical agglomerative cluster analysis of the 15.000 groups detected by RI algorithm in the last round of analysis: simple matching as similarity criterion and complete linkage as agglomerative method; (vi) for each cluster, consideration of just the group having the highest t_{CI} , as the representative group of that cluster; (vii) cut of the dendrogram with regard to the length of branches, and to the overall involvement of a large number of agents.

6. REGRESSION MODELS TO QUALITATIVELY ANALYZE THE SIZE OF THE RESULTING INTERSECTIONS

Having separately analyzed the considered aspects, i.e. *relational structures* and *shared processes and common functions,* by using specific community detection analyses, the final part of this work concerns the investigation of the presence of a specific pattern characterizing the subsets of agents that are simultaneously grouped together by CPM, IM and RI analysis. In Subsection 3.5 two statistics regarding the size of the obtained intersections have have been introduced. In this section, the suggested models are presented, with these statistics as dependent variables. In this way, the qualitative elements that characterize the size of the intersections are investigated.

6.1 THE LINEAR REGRESSION MODELS

The final analysis of this work aims to investigate what characterizes the size of the obtained intersections based the types of activities that the involved agents performed. More specifically, the technological domains of the projects in which the agents participated, are considered so to uncover relevant behaviors from a qualitative point of view[29]. In order to develop a qualitative analysis, the number of participations that the agents in the intersection realized, is not considered as a single variable. Given the availability of information regarding the technological domain of each project, and so of each participation (see Section 4), a set of continuous variables is considered. Each of these variables is generated as the count of the number of participations that the agents included in the intersection had, in one specific technological domain. The technological domains (in parentheses, the names of the corresponding variable as they are described in Equations 3-6 and in Tables 2-5) are: (i) Bio-technologiy (*BioTech*); (ii) Chemistry (*Chem*); (iii) Geo-thermal Sciences and Biomasses (*Geot_Biom*); (iv) Information and Communication Technologies (*ICT*); (v) Mechanics (*Mech*); (vi) Nano-technologies (*NanoTech*); (vii) New Materials (*NewMat*); (viii) Optoelectronic (*Optoel*); (ix) Multi-disciplines (*Multi*); (x) others (*Oth*).

The following linear regression models (Rawlings et al., 2006) aim to qualitatively investigate how the participations of agents in projects having specific technological domains, characterize the size of the detected intersections. The

[29] In the case study of Region Tuscany innovation policies, the 'activities' are participations in network projects and, therefore, they are co-participations. Therefore, not only the behavioral dimension of agents is considered from a qualitative point of view, but it refers exactly to activities corresponding to interactions.

independent variables are the number of participations (of the agents included in the intersection) in the corresponding technological domain. The two models (Equations 3-4) both consider all the aforementioned independent variables. The first model (Equation 3) uses s_1 (i.e. the cardinality of the intersection) as dependent variable, and the second model (Equation 4) considers s_2 (i.e. the ratio between the cardinality of the intersection and the cardinality of the correspondent union set) as dependent variable.

$$s_1 = \beta_0 + \beta_1(BioTech) + \beta_2(Chem) + \beta_3(Geot\_Biom) + \beta_4(ICT) + \beta_5(Mech) + \beta_6(NanoTech) + \beta_7(NewMat) + \beta_8(Optoel) + \beta_9(Multi) + \beta_{10}(Oth) \quad (3)$$

$$s_2 = \beta_0 + \beta_1(BioTech) + \beta_2(Chem) + \beta_3(Geot\_Biom) + \beta_4(ICT) + \beta_5(Mech) + \beta_6(NanoTech) + \beta_7(NewMat) + \beta_8(Optoel) + \beta_9(Multi) + \beta_{10}(Oth) \quad (4)$$

Based on the significance of the parameters (Tables 2-3), other two models are finally formed as follows, in order to better represent what can be associated to size of the obtained intersections:

$$s_1 = \beta_1(\text{ICT}) + \beta_2(NanoTech) + \beta_3(NewMat) + \beta_4(Optoel) + d(Excluded) \quad (5)$$

$$s_2 = \beta_1(\text{ICT}) + \beta_2(NanoTech) + \beta_3(NewMat) + \beta_4(\text{O pt oel}) + \text{d(E xcluded)} \quad (6)$$

where the independent variable *d(Excluded)* is a dummy variable that assumes value 1 if at least one agent (among those involved in the intersection) participated in at least one of the technological domains that remained excluded from this second set of regressions[30].

By implementing these analyses, the size of the detected intersections is described by counting the participations (that the included agents had in the policy cycle) in terms of their type. Hence, the information based on a qualitative categorization of the agents' activities is considered, in order to describe the size of the corresponding intersection. All the considered variables (dependent and independent) are unity-based normalized [0, 1], within each of the subsets of results used to validate the models. In the next Subsection, the implemented process of validation is described.

---

[30] The technological domains that remained excluded are: Bio-technology or Chemistry or Geothermal Sciences and Biomasses or Multi-disciplines or others. Only the technological domain of Mechanics is completely excluded from this second set of models, because of its high correlation with the technological domain of ICT.

6.2 COMBINATIONS OF PARTITIONS TO VALIDATE THE ANALYSIS

In order to validate the approach, the regression models presented in Equations 3-6, are implemented on different subsets of the obtained results. Since each partition is a set of communities, the previously detected communities are already naturally subsetted. Therefore, for each of the models, instead of forming a single regression model that takes into account all the possible combinations of three communities as previously described[31], the partitions are considered to develop separate analyses. From the first part of the work, two partitions are obtained with CPM, three partitions with IM and one with RI; thence six combinations can be considered and analyzed separately. These six combinations, presented also in Figure 2(c), are listed below. The labels of the combinations that are analyzed through the models and that are presented in Tables 2-5, are indicated in parentheses.

- CPM_{k12} - IM_{Mkv1} - $RI_{\delta LoA}$ (Comb.A)
- CPM_{k12} - IM^{AS}_{Mkv2} - $RI_{\delta LoA}$ (Comb.B)
- CPM_{k12} - IM^{DS}_{Mkv2} - $RI_{\delta LoA}$ (Comb.C)
- CPM_{k18} - IM_{Mkv1} - $RI_{\delta LoA}$ (Comb.D)
- CPM_{k18} - IM^{AS}_{Mkv2} - $RI_{\delta LoA}$ (Comb.E)
- CPM_{k18} - IM^{DS}_{Mkv2} - $RI_{\delta LoA}$ (Comb.F).

All the possible combinations of one CPM partition and of one IM partition and of the RI partition, are considered. Subsequently, within each of these combinations, the areas of co-existence of the three aspects are investigated by taking into account the communities that belong to the involved partitions. For instance, to determine the intersections among CPM_{k12}, IM^{AS}_{Mkv2} and RI_{dLoA}, all the possible intersections among (i) one of the communities from CPM_{k12} , (ii) and one of the communities from IM^{AS}_{Mkv2} , and (iii) one of the communities from RI_{dLoA}, are taken into account. All the resulting intersections are considered as units of analysis.

6.3 RESULTS

The results of the models described in Equations 3-6 are reported in Tables 2-5. In each table, each column corresponds to the implementation of the considered model to a specific combination of partitions, as previously described. For each model are presented: the coefficients of the considered variables, the t statistics and their level of significance, the number of observations (i.e. the number of

[31] The considered combinations are necessarily generated by one CPM community and one IM community and one RI community. See Subsection 3.5.

intersections), and the adjusted R^2 that is used to allow the comparison among the models and the evaluation of their efficacy.

Tab. 2: Results of the linear regression model described in Equation 3 (s_1 as dependent variable). Intersections obtained through different combinations of partitions (columns), obtained with different community detection methodologies, are considered as described in Subsection 6.2.

	Comb.A	Comb.B	Comb.C	Comb.D	Comb.E	Comb.F
	s_1	s_1	s_1	s_1	s_1	s_1
Oth	-0.149**	-0.364***	-0.343***	-0.0329	-0.318***	-0.269***
	(-2.71)	(-17.12)	(-15.33)	(-0.83)	(-10.34)	(-7.74)
BioTech	0.0509**	-0.0288	-0.00373	0.00807	-0.0911***	-0.0698**
	(2.63)	(-1.46)	(-0.17)	(0.22)	(-5.00)	(-2.95)
Chem	0.0365	0.0885***	0.0985***	0.0363	-0.00403	0.0114
	(1.00)	(4.67)	(4.85)	(0.68)	(-0.16)	(0.39)
Geot_Biom	-0.0562	-0.0660***	-0.0621***	-0.0166	-0.0520**	-0.0593**
	(-1.54)	(-3.82)	(-3.49)	(-0.25)	(-2.95)	(-2.76)
ICT	0.667***	0.956***	0.956***	0.656***	0.956***	0.942***
	(21.48)	(28.22)	(28.30)	(13.36)	(28.72)	(26.44)
Mech	0.243***	0.228***	0.233***	0.290***	0.263***	0.259***
	(6.33)	(8.14)	(8.25)	(4.21)	(9.57)	(7.88)
Multi	0.00604	0.0771***	0.0911***	0.00657	0.104***	0.122***
	(0.25)	(3.82)	(4.78)	(0.25)	(5.27)	(5.56)
NanoTech	0.331***	0.541***	0.518***	0.168***	0.323***	0.256***
	(8.78)	(18.80)	(20.24)	(4.76)	(13.66)	(10.56)
NewMat	-0.260***	-0.308***	-0.351***	-0.227***	-0.303***	-0.291***
	(-6.96)	(-11.60)	(-13.86)	(-4.29)	(-12.38)	(-9.44)
Optoel	0.288***	0.220***	0.212***	-0.0134	0.0583***	0.0501**
	(6.76)	(6.49)	(6.97)	(-0.38)	(3.47)	(2.69)
_cons	0.0103	-0.0441***	-0.0503***	0.00783	-0.0666***	-0.0676***
	(1.25)	(-8.63)	(-10.37)	(0.71)	(-12.19)	(-12.41)
N	404	1011	1067	160	472	489
adj. R^2	0.769	0.800	0.795	0.849	0.895	0.878

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In all the models, the variable representing the participations of agents in *ICT* projects presents the most significant coefficients, as it can be observed in Tables 2-3. The low p-values of the variable *ICT* indicate that the addition of this variable is meaningful for the interpretation of both the dependent variables s_1 (see Table 2) and s_2 (see Table 3). Therefore, independently from the combination of partitions, the role of ICT projects appears as determinant in characterizing the activities of agents that belong to intersections. The participations in projects whose technological domains are Mechanics (*variable Mech*) and Nanotechnologies (*variable NanoTech*) appear also to hold a significant role. In comparison to those

of the variable *ICT,* their coefficients have smaller magnitudes and are associated to marginally lower t-statistics. The participations in Optoelectronics projects (*variable Optoel*) emerges also as significant, with values that are almost large as those of participations in Mechanics (*variable Mech*). The negative coefficients of the variable describing the participations in the technological domain of New Materials (*variable NewMat*), which has significant values of the t-statistic, indicate that in its presence the size of the intersections decreases. Participations in this domain are not significant only when analyzing statistic s_2 in the combination made of partition CPM_{k12}, partition IM_{Mkv1} and partition $RI_{\delta LoA}$ (Table 3, column 'Comb.D'). The participations in projects having Geo-thermal Sciences and Biomasses (*variable Geot_Biom*) as technological domain also present negative coefficients. However, the values of the t statistic indicate that this variable is less significant.

Tab. 3: Results of the linear regression model described in Equation 4 (s_2 as dependent variable). Intersections obtained through different combinations of partitions (columns), obtained with different community detection methodologies, are considered as described in Subsection 6.2.

	Comb.A	Comb.B	Comb.C	Comb.D	Comb.E	Comb.F
	s_2	s_2	s_2	s_2	s_2	s_2
Oth	-0.180**	-0.309***	-0.318***	-0.0926	-0.275***	-0.230***
	(-3.22)	(-14.12)	(-13.56)	(-1.90)	(-9.19)	(-6.59)
BioTech	0.0686*	0.0273	0.0365	0.0450	-0.0591**	-0.0454
	(2.37)	(1.38)	(1.66)	(1.03)	(-3.07)	(-1.91)
Chem	0.0646	0.0855***	0.0846***	0.175*	-0.0125	-0.0120
	(1.22)	(4.45)	(3.96)	(2.22)	(-0.44)	(-0.34)
Geot_Biom	-0.105*	-0.0866***	-0.0863***	-0.145	-0.0605***	-0.0683**
	(-2.11)	(-4.63)	(-4.39)	(-1.65)	(-3.34)	(-2.94)
ICT	0.630***	0.748***	0.769***	0.617***	0.771***	0.792***
	(13.08)	(18.71)	(18.74)	(6.91)	(18.94)	(17.73)
Mech	0.0764	0.106**	0.114**	0.160	0.129***	0.150***
	(1.35)	(3.12)	(3.22)	(1.45)	(3.91)	(3.44)
Multi	-0.0508	0.0934**	0.0895**	-0.103*	0.134***	0.124***
	(-1.08)	(2.95)	(2.95)	(-2.59)	(4.67)	(4.09)
NanoTech	0.267***	0.413***	0.447***	0.128**	0.276***	0.227***
	(5.49)	(14.06)	(15.67)	(2.68)	(9.09)	(7.66)
NewMat	-0.227***	-0.265***	-0.318***	-0.120	-0.289***	-0.272***
	(-4.28)	(-9.26)	(-11.27)	(-1.48)	(-9.08)	(-7.43)
Optoel	0.222***	0.0722*	0.105***	-0.0529	0.0175	0.0125
	(3.77)	(2.17)	(3.34)	(-1.00)	(0.83)	(0.55)
_cons	0.116***	0.00705	0.00829	0.103***	-0.0193*	-0.0224**
	(9.21)	(1.06)	(1.21)	(5.60)	(-2.44)	(-2.85)
N	404	1011	1067	160	472	489
adj. R^2	0.447	0.601	0.565	0.647	0.779	0.731

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Since very similar emerging patterns are detected throughout all the considered combinations that validate the models defined in Equations 3-4, additional regressions are implemented. In order to provide a more global model, i.e. considering less explanatory variables, the models described in Equations 5-6 are developed. The statistics s_1 and s_2 are again considered as dependent variables, and the number of independent variables is reduced. Based on Tables 4-5, for most of the analyzed cases, the independent variables are significant for a significance level of 99.9%. The dummy variable *d(Excluded)* representing participations in projects with technological domains that are not directly considered in the models, appears as less significant. This justifies the development of these models. Moreover, the adjusted R^2 is higher for models in Tables 4-5, than in models in Tables 2-3. These results indicate that the more refined last two models describe better the dependent variables than the initially proposed models.

Tab. 4: Results of the linear regression model described in Equation 5 (s_1 as dependent variable). Different combinations of partitions (columns), obtained with different community detection methodologies, are considered as described in Subsection 6.2.

	Comb.A	Comb.B	Comb.C	Comb.D	Comb.E	Comb.F
	s_1	s_1	s_1	s_1	s_1	s_1
ICT	0.730***	0.997***	0.988***	0.816***	1.021***	0.989***
	(30.59)	(36.93)	(37.38)	(28.31)	(31.96)	(29.82)
NanoTech	0.272***	0.239***	0.272***	0.118***	0.0911***	0.0944***
	(8.24)	(9.97)	(11.72)	(4.21)	(5.02)	(5.01)
NewMat	-0.211***	-0.142***	-0.190***	-0.179***	-0.114***	-0.128***
	(-6.22)	(-6.18)	(-8.29)	(-3.75)	(-4.51)	(-4.86)
Optoel	0.297***	0.203***	0.201***	-0.0769**	-0.0643***	-0.0391**
	(7.72)	(6.11)	(7.81)	(-3.01)	(-3.59)	(-2.59)
d(Excluded)	0.0314***	-0.0328***	-0.0288***	0.0200	-0.0476***	-0.0400***
	(4.35)	(-6.05)	(-5.55)	(1.83)	(-7.26)	(-6.56)
N	404	1011	1067	160	472	489
adj. R^2	0.904	0.872	0.874	0.941	0.911	0.909

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The statistical analysis of the models for the aforementioned combinations (Tables 2-5) indicates that the suggested models represent pertinently the size of the intersections. Therefore, the size of the areas in which *structures* and *processes* and *functions* co-exist (represented by both the dependent variables) is qualitatively characterized by specific types of activities performed by the involved agents (represented by the independent variables). The results of this final part of the work, which are validated through the different subsets of the results of the first part of the present work (see Subsection 6.2), indicate that the entire methodological approach here developed, leads to the identification of meaningful non-random areas of the socio-economic system.

Tab. 5: Results of the linear regression model described in Equation 6 (s_2 as dependent variable). Different combinations of partitions (columns), obtained with different community detection methodologies, are considered as described in Subsection 6.2.

	Comb.A	Comb.B	Comb.C	Comb.D	Comb.E	Comb.F
	s_2	s_2	s_2	s_2	s_2	s_2
ICT	0.768***	0.815***	0.847***	0.887***	0.843***	0.871***
	(25.11)	(30.24)	(31.25)	(21.02)	(26.35)	(27.91)
NanoTech	0.249***	0.187***	0.251***	0.107**	0.0971***	0.101***
	(4.63)	(8.11)	(10.49)	(2.76)	(5.16)	(5.20)
NewMat	-0.287***	-0.156***	-0.217***	-0.180***	-0.139***	-0.142***
	(-5.42)	(-6.35)	(-9.03)	(-3.37)	(-5.43)	(-5.31)
Optoel	0.324***	0.126***	0.164***	-0.0743*	-0.0456**	-0.0374**
	(7.62)	(4.91)	(7.73)	(-2.29)	(-3.22)	(-2.71)
d(Excluded)	0.0792***	-0.000445	0.000206	0.0249	-0.0237**	-0.0271***
	(5.89)	(-0.07)	(0.03)	(1.34)	(-3.08)	(-3.73)
N	404	1011	1067	160	472	489
adj. R^2	0.807	0.808	0.799	0.885	0.881	0.868

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

7. CONCLUSIONS

The current work proposes a methodological solution to investigate socio-economic complex systems with innovative dynamics, by combining community detection analyses. Based on a specific theoretical background (Lane, 2011), innovation is considered as a phenomenon blooming from interactions. Three aspects are identified as crucial: (i) *relational structures,* (ii) *shared processes* and (iii) *common functions.* Since their presence is suggested to potentially favor interactions that generate innovations, each of these aspects is investigated with a specific community detection analysis. Clique Percolation Method (CPM) is implemented to investigate how agents established certain architectures of relationships (i.e. *relational structures*). Infomap algorithm (IM) is implemented to identify, through the simulation of flows, the exchanges of information that are associated with economic initiatives involving a plurality of agents (i.e. *shared processes*). The Relevance Index (RI) is implemented to investigate the integration of agents' behaviors over time, so as to unveil agents' similarity in terms of roles and/or goals (i.e. common functions).

Since the theoretical approach defines the joint co-existence of *structures, processes,* and *functions* as a crucial condition for the genesis of innovation, intersections among communities detected by different methodologies are finally computed, as they are intended to represent areas of entire intersection among all

the aforementioned aspects. In this scope, groups of agents simultaneously grouped in (i) one CPM community, and (ii) one IM community, and (iii) one RI community, are considered. The size of these intersections is investigated by computing two statistics, namely s_1 (i.e. the cardinality of the intersection) and s_2 (i.e. the ratio between the cardinality of the intersection and the cardinality of the correspondent union set), that are used as dependent variables in linear regression models. The aim of these models is to qualitatively analyze which types of interactions are developed by agents included in these areas of intersections. Hence, a relation between the size of the areas where *structures-processes-functions* co-exist, and the development of specific types of activities by involved agents, is addressed.

Being validated with different combinations of subsets of the output of the community detection analyses, the four suggested regression models uncover regular elements that demonstrate that this relation is not random. In particular, the recurrence of the values of the coefficients of the variables for the different combinations of partitions, and the good fitness to data (i.e. high values of R^2), suggest that the models are efficiently representing the size of the intersections. Therefore, the assumed relation between the size of the intersections and the development of specific types of interactions, is confirmed. However, the objective of this final part of the analysis is neither to prove the validity of a specific model, nor to find correspondence with elements addressed as relevant in literature. The main aim of the proposed regression analysis is to prove that the developed methodological approach leads to results in which regular elements emerge.

A lack of randomness in the relation between the size of the intersections determined by the whole developed methodological approach, and the types of activities performed by involved agents, is hence determined. Therefore, the combination of community detection analyses and the final investigation of intersections (provided by the use of different methodologies), enable the uncovering of areas of the system that deserve to be further considered. In addition, also other areas of non-entire intersections, i.e. those presented in Figure 2(d) but not highlighted in Figure 2(e), should be further assessed. Analyses regarding the evolution of these areas and their role in the reproductive dynamic of the corresponding system should be developed. The comprehension of which are the aspects that agents are able to develop through their interactions, has to be associated with the comprehension of how agents survive, evolve and reproduce through a mutual relation with the environment in which they live. Through these insights, objectives can be defined and, through adequate normative interventions, the evolution of the system can be driven towards specific directions.

ACKNOWLEDGMENTS

I gratefully acknowledge the kind support of the following for the development of this work: Giuditta De Prato, Enrico Giovannetti, Simone Righi, Margherita Russo, Sofia Samoili, Roberto Serra, Marco Villani, Maria Prosperina Vitale and anonymous reviewers.

REFERENCES

Amendola, M. and Gaffard, J.-L. (1988). *The Innovative Choice.* Basil Blackwell, Oxford, UK.

Amendola, M. and Gaffard, J.-L. (1998). *Out of Equilibrium.* Clarendon Press, Oxford, UK.

Breschi, S. and Malerba, F. (1997). Sectoral innovation systems: technological regimes, schumpeterian dynamics, and spatial boundaries. *Systems of Innovation: Technologies, Institutions and Organizations,* pages 130–156.

Burt, R. (1995). *Structural Holes.* Harvard University Press, Cambridge, Massachussets, US.

Caloffi, A., Rossi, F. and Russo, M. (2015). The emergence of intermediary organizations: A network-based approach to the design of innovation policies. In Cairney, P. and R. Geyer, editors, *Handbook* On Complexity And Public Policy, pages 314–331. Edward Elgar, Cheltenham, UK.

Coase, R.H. (1937). The nature of the firm. In *Economica,* 4(16): 386–405.

Dosi, G., Freeman, C., Nelson, R., Silverberg, G. and Soete, L. (1988). *Technical Change and Economic Theory,* volume 988. Pinter Publishers, London, UK.

Etzkowitz, H. and Leydesdorff, L. (2000). The dynamics of innovation: from national systems and mode 2 to a triple helix of university–industry–government relations. *Research Policy,* 29(2): 109–123.

Filisetti, A., Villani, M., Roli, A., Fiorucci, M. and Serra, R. (2015). Exploring the organisation of complex systems through the dynamical interactions among their relevant subsets. In *Proceedings of the European Conference on Artificial Life,* pages 286–293. MIT Press.

Fortunato, S. (2010). Community detection in graphs. In *Physics Reports,* 486(3): 75–174.

Freeman, C. (1991). Networks of innovators: a synthesis of research issues. In *Research Policy,* 20(5): 499–514.

Georgescu-Roegen, N. (1971). *The Entropy Law and the Economic Process.* Harvard University Press, Cambridge, Massachussets, US.

Hicks, J.R. (1973). *Capital and Time: A Neo-Austrian Theory.* Clarendon Press, Oxford, UK.

Hidalgo, C. (2015). *Why Information Grows: The Evolution of Order, from Atoms to Economies.* Basic Books, New York, US.

Huffman, D.A. (1952). A method for the construction of minimum-redundancy codes. In *Proceedings of the IRE,* 40(9): 1098–1101.

Lane, D.A. (2011). Complexity and innovation dynamics. In *Handbook on the Economic Complexity of Technological Change,* volume 63. Edward Elgar, Cheltenham, UK.

Lane, D.A. and Maxfield, R.R. (1997). Foresight, complexity, and strategy. In Arthur, W.B., Durlauf, S.N. and Lane, D.A., editors, *The Economy as an Evolving Complex System II.* Westview Press, Reading, Massachussets, US.

Lane, D.A. and Maxfield, R.R. (2005). Ontological uncertainty and innovation. In *Journal of Evolutionary Economics,* 15(1): 3–50.

Lundvall, B.-A., Dosi, G., and Freeman, C. (1988). Innovation as an interactive process: From user-producer interaction to the national system of innovation. In Dosi, G., Freeman, C., Nelson, R., Silverberg, G., and Soete, L., editors, *Technology and economic theory.* Pinter Publishers, London, UK.

Malerba, F. and Orsenigo, L. (1997). Technological regimes and sectoral patterns of innovative activities. In *Industrial and Corporate Change,* 6(1): 83–118.

Marx, K. (1867). *Das kapital: Kritik der Politischen Ökonomie.* Verlag von Otto Meisner, Hamburg, Germany.

Mowery, D.C. and Teece, D.J. (1996). Strategic alliances and industrial research. *Engines of Innovation: US Industrial Research at the End of an Era,* pages 111–129.

Nelson, R.R. (1993). *National Innovation Systems: a Comparative Analysis.* Oxford University Press, New York, US.

Palla, G., Derényi, I. and Vicsek, T. (2005). Clique percolation in random net-works. In *Nature,* 435(7043): 814–818.

Picchio, A. (1992). *Social Reproduction: the Political Economy of the Labour Market.* Cambridge University Press, Cambridge, UK.

Quesnay, F. (1758). Tableau économique. Manuscript.

Rawlings, J., Pantula, S. and Dickey, D. (2006). *Applied Regression Analysis: A Research Tool.* Springer Texts in Statistics. Springer, New York, US.

Righi, R., Roli, A., Russo, M., Serra, R. and Villani, M. (2017). New paths for the application of DCI in social sciences: Theoretical issues regarding an empirical analysis. In Rossi, F., Mavelli, F., Strano, P. and Caivano, D. editors, *Advances in Artificial Life, Evolutionary Computation, and Systems Chemistry,* pages 42–52. Springer International Publishing.

Roli, A., Villani, M., Caprari, R. and Serra, R. (2017). Identifying critical states through the relevance index. In *Entropy,* 19(2): 73–88.

Rosvall, M. and Bergstrom, C.T. (2008). Maps of random walks on complex networks reveal community structure. In *Proceedings of the National Academy of Sciences,* 105(4): 1118–1123.

Rosvall, M. and Bergstrom, C.T. (2010). Mapping change in large networks. In *PloS One,* 5(1): e8694.

Rosvall, M., Esquivel, A.V., Lancichinetti, A., West, J.D. and Lambiotte, R. (2014). Memory in network flows and its effects on spreading dynamics and community detection. In *Nature Communications. 5, 4630.*

Russo, M. (2000). Complementary innovations and generative relationships: an ethnographic study. In *Economics of Innovation and New Technology,* 9(6): 517–558.

Sani, L., Amoretti, M., Vicari, E., Mordonini, M., Pecori, R., Roli, A., Villani, M., Cagnoni, S. and Serra, R. (2016). Efficient search of relevant structures in complex systems. In Adorni, G., Cagnoni, S., Gori, M. and Maratea, M., editors, *AI*IA 2016 Advances in Artificial Intelligence,* pages 35–48. Springer International Publishing.

Saxenian, A. (1994). Regional networks: industrial adaptation in Silicon Valley and route 128. Cityscape: *A Journal of Policy Development and Research,* 2(2): 41–60.

Sen, A. (1985). *Commodities and Capabilities.* Oxford University Press, Oxford, UK.

Sen, A. (1987). *On Ethics and Economics.* The Royer Lectures. Wiley, Oxford, UK.

Shannon, C.E. (1948). A mathematical theory of communication. In *The Bell System Technical Journal,* 27: 623–656.

Smith, A. (1761). *The Theory of Moral Sentiments.* A. Millar, London, UK.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations.* W. Strahan and T. Cadell., London, UK.

Tononi, G., McIntosh, A.R., Russell, D.P. and Edelman, G.M. (1998). Functional clustering: Identifying strongly interactive brain regions in neuroimaging data. In *NeuroImage,* 7(2): 133–149.

Tononi, G., Sporns, O. and Edelman, G.M. (1994). A measure for brain complexity: Relating functional segregation and integration in the nervous system. In *Proceedings of the National Academy of Sciences,* 91(11): 5033–5037.

Villani, M., Filisetti, A., Benedettini, S., Roli, A., Lane, D. and Serra, R. (2013). The detection of intermediate-level emergent structures and patterns. In *Advances in Artificial Life: ECAL 2013,* pages 372–378. MIT Press.