New Paths for the Application of DCI in Social Sciences: Theoretical Issues Regarding an Empirical Analysis

Riccardo Righi[1(✉)], Andrea Roli[2], Margherita Russo[1], Roberto Serra[3], and Marco Villani[3]

[1] Department of Economics, University of Modena and Reggio, Emilia, Italy
{riccardo.righi,margherita.russo}@unimore.it
[2] Department of Computer Science and Engineering, University of Bologna, Bologna, Italy
andrea.roli@unibo.it
[3] Department of Physics, Informatics and Mathematics, University of Modena and Reggio, Emilia, Italy
{rserra,marco.villani}@unimore.it

Abstract. Starting from the conceptualization of 'Cluster Index' (CI), Villani *et al.* [16,17] implemented the 'Dynamic Cluster Index' (DCI), an algorithm to perform the detection of subsets of agents characterized by patterns of activity that can be considered as integrated over time. DCI methodology makes possible to shift the attention into a new dimension of groups of agents (i.e. communities of agents): the presence of a common function characterizing their actions. In this paper we discuss the implications of the use in the domain of social sciences of this methodology, up to now mainly applied in natural sciences. Developing our considerations thanks to an empirical analysis, we discuss the theoretical implications of its application in such a different field.

1 Introduction

Taking advantage of two information theory concepts (integration and mutual information), Giulio Tononi introduced a new concept, the 'cluster index' (CI) [12–15], and demonstrated that neurons with integrated profiles of activity over time (i) have similar functions and (ii) have a location that is independent from anatomical proximity. Following this pioneering contribution, Villani and coauthors [16,17] developed an algorithm for the detection of subsets of agents introducing the comparison between the CI of an observed subset and the CI of a homogeneous system. The resulting algorithm, named by the authors 'Dynamic Cluster Index' (DCI), is able to produce a final ranking of all possible subsets that can be considered in any initial set. So far, DCI has been tested in research areas of artificial network models, of catalytic reaction networks and of biological gene regulatory systems [16,17], giving a contribution to the problem of identifying emergent meso-level structures [3].

F. Rossi et al. (Eds.): WIVACE 2016, CCIS 708, pp. 42–52, 2017.
DOI: 10.1007/978-3-319-57711-1_4

The implementation of the DCI algorithm opens new paths for addressing socio-economic problems regarding the analysis of group of agents. In social sciences, the detection and the analysis of communities typically are performed through the consideration of similar characteristics of agents, or through the analysis of the observed network structure. Indeed, DCI methodology makes possible to shift the attention into a new dimension of organizations of agents: the presence of a common function characterizing their actions. Developing our considerations thanks to an empirical analysis, in this paper we discuss the theoretical implications of the use of this methodology in the domain of social sciences.

The paper is structured as follows. Section 2 proposes an overview of the theoretical elements of the CI proposed by Tononi *et al.* [12–15] and of the DCI as proposed by Villani *et al.* [16,17]. Section 3 presents the issue addressed in the case study: the evaluation of the network innovation regional policy implemented by Tuscany Region (Italy) in the programming period 2000–2006. Section 4 discusses the advantages of applying DCI in a context where the application of Complex Network modeling of community detection comes up against the absence of stepwise processes of formation/dissolution of relational structures. Section 5 focuses on the analytical problem of defining what the "activity" of an agent is. In Sect. 6, theoretical considerations regarding the application of DCI in a socio-economic context of analysis are illustrated. Section 7 underlines the potentiality of DCI analysis to investigate unobserved relations. Section 8 concludes summarizing the investigation of functional communities (group of agents that share a common function) in the landscape of community detection techniques, and highlights the potentialities of the application of DCI in socio-economic analyses aiming at detecting emerging functional communities.

2 DCI Analysis for the Detection of Functional Groups of Agents

Dynamic Cluster Index analysis (DCI) takes its origin from the neurological studies of Giulio Tononi in the 90's. Tononi supposed that neurons with similar functions show high level of coordination in their behaviors over time, independently from being, or not, situated in the same brain region[1]. 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" [12]. Thus, these functional clusters are made up of interactive neurons that produce an internal exchange of information (among neurons belonging to the same group) stronger

[1] 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 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.

than the exchange of information that the same neurons have with the rest of the system. The identification of these groups of neurons was realized by making use of two information theory concepts derived from the Shannon entropy: integration (I) and mutual information (MI). Taking advantage of these measurements a new concept was introduced [14]: the cluster index (CI). Formally, the CI of a subset X is written as follows

$$CI(X) = I(X)/MI(X, U \backslash X) \tag{1}$$

where X is the *j-th* subset of the whole system U, and is made up of k elements. Thanks to the CI, evidences of neurons that could be considered to belong to specific functional sub-systems, even if they do not participate in the same cerebral area, were found [14]. These studies demonstrated that neurons showing integrated profiles of activity over time (grouped together thanks to the analysis of CI) have (i) similar functions in the brain, and (ii) not necessarily show anatomical proximity.

Since integration and mutual information values depend on the size of the subsystem that is under analysis, in order to normalize them it is possible to make use of a so-called homogeneous system where the variables do not have correlation[2] [14,16,17]. Finally, the level of significance of the normalized CI (calculated as a statistical distance of the normalized CI, or CI', of the considered subset from the average CI of a subset having the same size extracted from the homogeneous system) is the value according to which the final ranking of all possible subsets is produced [14]:

$$CI'(X) = \frac{I(X)}{\langle I_h \rangle} \Big/ \frac{M(X, U \backslash X)}{\langle M_h \rangle} \tag{2}$$

$$t_{ci} = \frac{CI'(X) - \langle CI'_h \rangle}{\sigma(CI'_h)} \tag{3}$$

where $\langle \mathrm{I_h} \rangle$ and $\langle \mathrm{M_h} \rangle$ indicate respectively the average integration of subsets of dimension k belonging to the homogeneous system and the average mutual information of these subsets with the remaining part of the homogeneous system. $\langle \mathrm{CI'_h} \rangle$ and $\sigma(\mathrm{CI'_h})$, respectively the mean and the standard deviation of normalized cluster indices of subsets that have the same size of X and that belong to the homogeneous system, are used to compute the statistical index $\mathrm{t_{ci}}$.

Following these studies, Villani and co-authors borrowed the concept of CI and $\mathrm{t_{ci}}$ introducing it in research areas of artificial network models, of catalytic reaction networks and of biological gene regulatory systems, giving a contribution to the problem of identifying emergent meso-level structures [8,16,17]. Moreover, since an exhaustive computation of this statistic ($\mathrm{t_{ci}}$) is possible only in small artificially designed networks, like those that were initially used to test

[2] A homogeneous system is a system having the same number of variables of the system to which it is referred; each variable has a random generated behavior in accordance with the probability of the states it assumes in the reference system.

the efficacy of the method [3,16,17], Villani and co-authors overcome the problem of computational duration in bigger initial set introducing a heuristic investigation in the algorithm [3]. The creation and the implementation of the DCI algorithm opened new path for analyses of community detection. The process of investigation made possible by DCI, allows researcher to look for groups characterized by levels of behavioral integration that, being significantly far from randomness, reveal the presence of at least one specific function jointly pursued by all the involved members. So far, the detection of groups (or communities) has been typically performed by focusing on similarity of agents' characteristics, or through the analysis of the observed network structure. With DCI methodology it is possible to shift the attention into a new dimension of organizations of agents. Since low levels of entropy are determined by the repetition of specific combinations of the statuses of a multiplicity of individuals over time, the emergence of a dynamic pattern unveils the alignment of the actions of these individuals towards a common function.

3 The Case Study: A Regional Programme to Support Innovation Networks

To interpret the application of DCI in social sciences, in this section we present an empirical analysis on a regional programme implemented by Tuscany Region (Italy) in the period 2000–06, aimed at supporting innovation networks. The programme sustained the development of innovation processes by fostering interactions between local agents (enterprises, universities, public research centers, local government institutions, service centers, etc.). The rationale of those policies is traced back in the need to overcome the difficulties of an industrial structure characterized by small and medium size enterprises in traditional sectors (such as textile, fashion, marble), generally not well linked to research activities. The various policy measures allowed the granting of funds exclusively to projects promoted by group of agents. Based on complex network analysis of innovation processes [5–7], that policy has been analysed by Caloffi, Rossi and Russo who highlighted the ontology of the programme [11], created an original data base with information on the agents participating to the programme and investigated the characteristics of the agents participating in the network-projects [1,2,9,10]. Since the goal of the policy programme was to favor collaborations in order to stimulate the flourishing of innovation, a key research question is to assess whether these policies were contributing to the formation of communities of innovative agents.

Started in 2002 (ending in 2008), the programme 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 addition, each 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 asked 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.

Another key feature was that agents could participate in more than one project (irrespective of the waves in which the projects were submitted) and they participate repeatedly with different partners. This means that every observed partnership could be made up of a different combination of agents and this element, added the those described above, increases the difficulties of grasping the network dynamics emerged over time. Looking at Fig. 1, that presents agents' participations over the nine waves (each agent keeps always the same coordinates across the nine graphs, and is represented only if in the correspondent wave it was active), it is possible to immediately capture how all the features described above produced a discontinuous evolution of the network. In such a picture, it seemed not appropriate to study the flourishing of communities looking at the stepwise creation of network frameworks. The degree of formation and of dissolution of connections was so high that brought to a situation of intense discontinuity over time.

4 DCI Analysis and Discontinue Network Dynamics

The peculiarities described in the previous paragraph strongly affected the possibility to study the detection of communities by taking into account the stepwise formation of relational structures among agents. Moreover, since the specific objective of our analysis is to investigate the presence of agents having a common function (and in this case study the common functions are likely to be related with the participation in projects and with the development of innovations), the analysis could not start from information on project-networks. In fact, following Tononi *et al.* [14], the presence of a functional groups is associated with the emergence of specific patterns of interactions, more than with the progressive establishment of relational community structures. We are not looking for the presence of a connective architecture that in some instant in time could reveal the presence of a community: we are looking for dynamics of interactions that reveal an alignment of agents and, consequently, that imply the joint pursuit of a common goal. Thus, rather than considering the formation/dissolution of the established connective structures, the idea was to focus the attention into how agents behaved over time. These considerations led to the formulation of a specific research question: *how is it possible to investigate the presence of functional communities, in a context in which the involved agents interact over time, but through networks' configurations that are discontinously changing over time?*

To answer this question, DCI analysis seemed to be appropriate. Any kind of relational information (topology of the project-network configuration) was

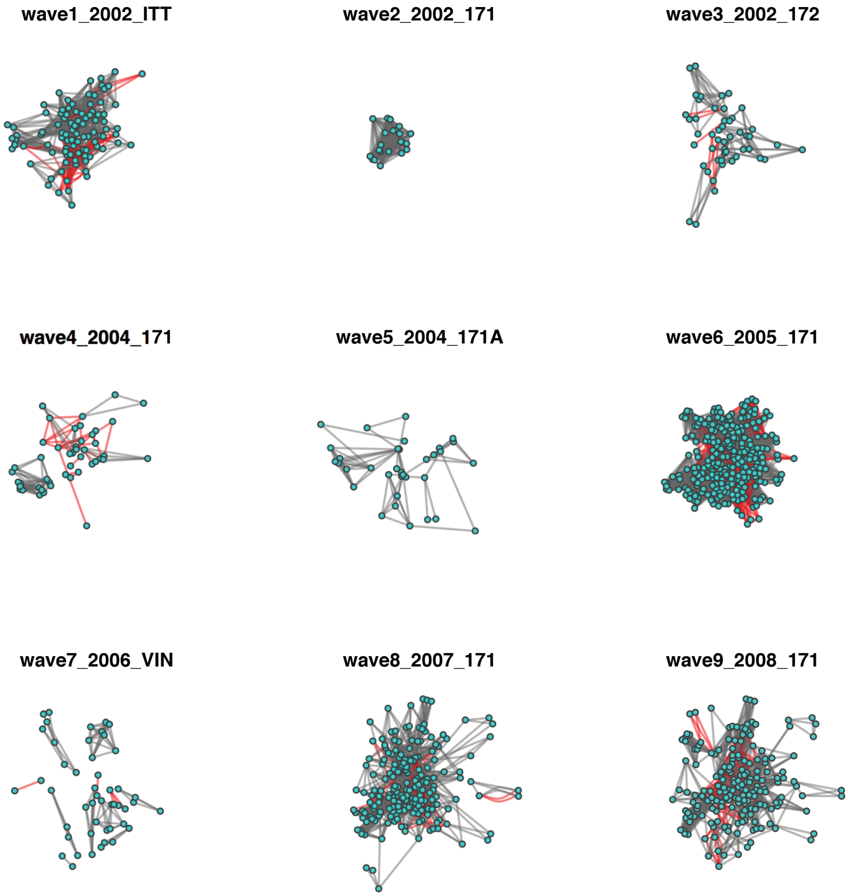

Fig. 1. Graphs representing the agents that took part in each of the nine waves. Each wave is labelled with a progressive number reflecting the overall chronological order, with the year in which it began, and with a code associated with the specific kind of the policies promoted. To every node is attributed the same position in all the graphs. Nodes 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. Source: our elaboration on Region Tuscany Network Policies (2000–06). Figure generated using `R (igraph package)`.

abandoned and the attention was centered on the integration of agents' behaviors over time. The idea was to observe what agents did over the period of the considered public policies.

5 The Application of DCI in a Socio-economic Context of Analysis

While up to now the motivations that led to the application of DCI analysis in a context of policy evaluation have been described, there are some theoretical implications that need to be taken into account. Moving to a socio-economic context of analysis necessarily involves three different considerations. First of all, it is more difficult to isolate the system under analysis, and since agents may be influenced by external elements, a high degree of openness can compromise the analysis enhancing the difficulty to investigate the dynamic of agents' subsystems that are the focus of the analysis. Secondly, while in the context of artificial and chemical systems initial conditions can be easily established, making it relatively easy to understand the relation between them and the behaviors of agents, this is much harder to do in socio-economic systems. Finally, there is also an ontological issue that must be considered: interactions of socio-economic agents may reveal behavioral patterns, but agents do not necessarily follow deterministic rules of interaction.

Without having the purpose to investigate the specific and detailed functions of subsets of agents, the aim of applying DCI in social sciences is to evaluate the presence of common objectives which, if they exist, are likely to imply the presence of information embodied in integrated activities. To have a shared function (as it is supposed that members of a community have) implies that involved agents behave with some kind of physical order that help them to reach together their common aim. The contraposition is between random behaviors, which do not give sense to a group action, and the physical order through which agents act to perform a shared function [4]. The information that the integration of behaviors contains is exactly what DCI takes into account.

6 The Data Structure and the Definition of Agents' Activity

The most important aspect that we had to face in order to apply DCI analysis to the case study regarded the definition of the agents' activity statuses. Since the available information regarded the starting and the ending date of agents' participations in the projects, we defined that each agent had to be considered active in those moments in which it was participating in at least one project. Working on these dates, it was possible to define a complete behavioral profile for the agents involved in the six years of the policy programme: in every instant it was possible to observe which agents were active and which were not. First of all boolean variables were defined, in accordance with the participation of agents in at least one project. Moreover, since information about the number of projects in which every agent was active in each single instant was available, it was possible to define a second series of variables describing the variation of the levels of activity. These variables assume four different values that correspond to four different situations:

- the agent is not participating in any project (no activity);
- 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);
- 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);
- 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).

With this second series of variables, we create of a model that takes into account a second order Markov condition. Agents' activity is not described for what is in each instant, but for what is in the present conditioned by what was in its nearest past. This more detailed description of the agents' behavioral dynamics allows evaluating entropy measurements with respect to how behaviors were changing. By doing so, more information is introduced in the model.

To conclude this section, it is important to remark the fact that, even if information regarding participation in projects was the only one available, it was the best we could ask for. As it would be clearer in the next section, policies contributed in the rising of other kinds of activities and other kinds of interactions among agents. Nevertheless, what the measures of policy fostered them to do was to design project-networks. Thus, the participation in projects, with regard to the specific context of analysis, has to be considered the ultimate kind of activity to which they tended, and so the most relevant among all.

7 DCI Analysis and Unobserved Relations Among Agents

Finally, moving to the conclusion of this contribution, a last theoretical point regarding the application of DCI in such a context of analysis has to be introduced. As it has been described above, DCI analysis allows researchers to investigate the presence of communities of agents without considering any kind of information about the topology of the network, a specific feature that assumes a particular meaning in the case study presented here. Since the considered policy measures financed exclusively project-networks, an immediate solution to a problem of community detection could be to focus on the structure of common participations. Of course, in the analysis of the considered public policy these relations represent the crucial types of interactions but, nevertheless, they represent only one type among all those that during the entire policy programme could have occurred. Thus, a crucial question arises: *is it not possible that functional groups could have bloomed not only through common participations in projects?*

The answer is affirmative. It is likely to admit that during the considered period of time agents did not have relations among themselves exclusively on the basis of common participations in projects and, consequently, it is not unreasonable to think that other unobserved relations could have been important in agents' decisions about final participations in regional public policies. 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 communities. Obviously, from the point of view of our analysis, it would have been incredibly interesting to consider and to study these relations, but they remained unobserved. To focus the attention on agents' participations in the policies means to consider their ultimate activity: that kind of activity that is the final result of all other interactions that occurred. Thus, it is possible to assert that the application of DCI methodology, even if without taking into account network information, opens the opportunity to embed in the result the whole set of relations that were active during the period under observation.

8 Conclusion

DCI analysis shows several theoretical aspects that encouraged its application in the domain of social sciences. It has been clarified that the motivation for which it was developed opens news possibilities for researchers: the investigation of functional communities (group of agents whose patterns of activity, being far from randomness, reveal the presence of a common function) is something new in the landscape of community detection methodologies. It has been described that in a context of analysis in which available relational data show a strong discontinuity over time, it would have been difficult to look at the stepwise formation of specific network structures, while DCI does not require this kind of condition. It has been remarked that, in a socio-economic analysis, the intrinsic nature of the context and of the agents considered do not affect the possibility to apply DCI to investigate the presence of integrated behaviors. Then, some considerations about the definition of the activity status have been made. In particular, a model that takes into account a second order Markov condition has been described, and the importance of the kind of activity that are considered was explained. Finally, it has been highlighted that DCI allows researchers to make considerations upon the complete set of relations that occurred in the considered period, without necessarily constraining the analysis to those that were observable. This final point is crucial since relations that remain unobserved could have had a role in the process of the emerging of functional communities.

With this work we explained the theoretical considerations that open the possibility to apply DCI analysis in socio-economic contexts. It is important to remark that depending on the specific application (and on the availability of data), the results of the DCI should be integrated with the results of the application of other kinds of community detection algorithms and with analyses of the network structure. In particular, the comparison between network communities and DCI functional groups should give helpful insights in socio-economic complex systems.

References

1. Caloffi, A., Rossi, F., Russo, M.: The emergence of intermediary organizations: a network-based approach to the design of innovation policies. In: Handbook On Complexity And Public Policy, pp. 314–331. Edward Elgar Publishing, Cheltenham GBR (2015). ISBN: 978-1-78254-951-2
2. Caloffi, A., Rossi, F., Russo, M.: What makes SMEs more likely to collaborate? Analysing the role of regional innovation policy. Eur. Plan. Stud. **23**(7), 1245–1264 (2015). http://dx.doi.org/10.1080/09654313.2014.919250
3. Filisetti, A., Villani, M., Roli, A., Fiorucci, M., Serra, R.: Exploring the organisation of complex systems through the dynamical interactions among their relevant subsets. In: Proceedings of the European Conference on Articial Life 2015 (ECAL 2015), vol. 13, pp. 286–293 (2016). https://mitpress.mit.edu/sites/default/files/titles/content/ecal2015/ch054.html
4. Hidalgo, C.: Why Information Grows: The Evolution of Order, from Atoms to Economies. Basic Books, New York (2015)
5. Lane, D.A.: Complexity and innovation dynamics. In: Handbook on the Economic Complexity of Technological Change. Edward Elgar Publishing, Cheltenham (2011). ISBN: 978-0-85793-037-8
6. Lane, D.A., Maxfield, R.R.: Foresight, complexity, and strategy. In: The Economy as an Evolving Complex System II. Westview Press (1997). ISBN: 978-0-201-32823-3
7. Lane, D.A., Maxfield, R.R.: Ontological uncertainty and innovation. J. Evol. Econ. **15**(1), 3–50 (2005). doi:10.1007/s00191-004-0227-7
8. Roli, A., Villani, M., Caprari, R., Serra, R.: Identifying critical states through the relevance index. Entropy **19**(2), 73 (2017). Special issue: Complexity, Criticality and Computation, MDPI AG, Basel, Switzerland
9. Rossi, F., Caloffi, A., Russo, M.: Networked by design: can policy requirements influence organisations' networking behaviour? Technol. Forecast. Soc. Change **105**, 203–214 (2016). doi:10.1016/j.techfore.2016.01.004
10. Russo, M., Caloffi, A., Rossi, F.: Evaluating the performance of innovation intermediaries: insights from the experience of Tuscany's innovation poles. In: Plattform Forschungs- Und Technologieevaluierung, vol. 41, pp. 15–25 (2015). ISSN: 1726–6629
11. Russo, M., Rossi, F.: Cooperation networks and innovation. A complex system perspective to the analysis and evaluation of a EU regional innovation policy programme. Evaluation **15**, 75–100 (2009). doi:10.1177/1356389008097872
12. Tononi, G., McIntosh, A.R., Russell, D.P., Edelman, G.M.: Functional clustering: identifying strongly interactive brain regions in neuroimaging data. NeuroImage **7**(2), 133–149 (1998). doi:10.1006/nimg.1997.0313
13. Tononi, G., Sporns, O., Edelman, G.M.: A complexity measure for selective matching of signals by the brain. Proc. Natl. Acad. Sci. USA **93**(8), 3422–3427 (1996). ISSN: 0027–8424
14. Tononi, G., Sporns, O., Edelman, G.M.: A measure for brain complexity: relating functional segregation and integration in the nervous system. Proc. Natl. Acad. Sci. USA **91**(11), 5033–5037 (1994). http://www.jstor.org/stable/2364906. ISSN: 0027–8424
15. Tononi, G., Sporns, O., Edelman, G.M.: Measures of degeneracy and redundancy in biological networks. Proc. Natl. Acad. Sci. **96**(6), 3257–3262 (1999). http://www.pnas.org/content/96/6/3257. ISSN: 0027–8424,1091–6490

16. Villani, M., Benedettini, S., Roli, A., Lane, D., Poli, I., Serra, R.: Identifying emergent dynamical structures in network models. In: Bassis, S., Esposito, A., Morabito, F.C. (eds.) Recent Advances of Neural Network Models and Applications. Smart Innovation, Systems and Technologies, vol. 26, pp. 3–13. Springer, Cham (2014). doi:10.1007/978-3-319-04129-2_1
17. Villani, M., Filisetti, A., Benedettini, S., Roli, A., Lane, D., Serra, R.: The detection of intermediate-level emergent structures and patterns. In: ECAL, pp. 372–378 (2013)