

Multiple Factor Analysis to Visually Explore Collaboration Structures: the Case of Technological Districts

L'Analisi Fattoriale Multipla per la visualizzazione delle strutture di collaborazione nei Distretti Tecnologici

Maria Rosaria D'Esposito and Domenico De Stefano and Giancarlo Ragozini

Abstract In the present work we propose to analyze, through Multiple Factorial Analysis (MFA), the relational structure embedded in collaboration networks emerging in Technological Districts (TD) and observed across time occasions. In our opinion, MFA allows to deal with complex relational structures that lead to highly heterogeneous networks. We show, through a case study, how the solutions provided by the MFA can be interpreted in a natural and suitable way in the relational setting. They can furnish valuable information about the strength and typology of the collaboration structure and its evolution. As case study we focus on the analysis of a TD located in South Italy.

Abstract Nel presente lavoro ci si propone di analizzare, mediante l'Analisi Fattoriale Multipla (AFM), l'evoluzione temporale della struttura delle reti di collaborazione nei Distretti Tecnologici (DT) italiani. L'AFM consente, infatti a nostro avviso, di analizzare dati relazionali complessi che danno luogo a reti di natura altamente eterogenea. Attraverso un caso studio, si mostrerà come le soluzioni fattoriali ottenute mediante l'AFM possono essere interpretate in maniera opportuna dal punto di vista relazionale e come forniscano informazioni utili circa la natura delle relazioni attivate tra i membri del DT. Come caso studio ci si riferisce ad un DT nel sud Italia.

Key words: Technological districts, Collaboration networks, Two-mode networks, Multiple Factorial Analysis, Time-varying networks

Maria Rosaria D'Esposito

Dipartimento di Scienze Economiche e Statistiche, Università di Salerno, e-mail: mdesposi@unisa.it

Domenico De Stefano

Dipartimento di Economia, Università Ca' Foscari di Venezia e-mail: domenico.destefano@unive.it

Giancarlo Ragozini

Dipartimento di Scienze Politiche, Università di Napoli Federico II e-mail: giragoz@unina.it

1 Introduction

Collaboration among private companies, institutions, and public research organizations is a topic of growing importance in the agendas of both research and development (R&D) policymakers and governmental administrators.

In the last decade, in Italy the efforts have been directed in promoting firms innovation capability through their systemic aggregations to foster R&D activities on key technologies. To this aim, the Italian government, through the Minister of University and Research, MIUR, has created the so-called Technological Districts (TDs) in some carefully chosen geographic locations in the national territory ¹. TDs may be defined as geographical concentration of interconnected companies and associated institutions including end-producers, universities, research laboratories, and service providers, all focused on a specialized area of economic activity.

Most policies are based, either in the design and decision phase or in the implementation phase, on a network of agents who concur in the policies [8] to create a critical mass of firms, research labs and universities. The main interest here is the analysis of relationships embedded in collaboration networks that occur among these entities and its evolution over time. For this reason, Social Network Analysis (SNA) techniques can be fruitfully used to measure the extent of occurrence, outspread and stabilization of collaboration relationships [2].

In order to apply SNA techniques to monitor TDs collaboration networks, it is worth to consider that these latter are complex structures in which each collaboration tie can be observed across time occasions that span from the birth of the analyzed TD until its current state. There are several possibilities to measure collaboration in terms of relational ties, such as the joint participation to research projects, the co-patenting, etc. These latter give rise to two-mode networks (actors-by-events), and when considering TDs' collaboration structures we observe a special case of complex networks, namely time-varying two-mode networks (actors-by-events under different types of relationships across time occasions). Such complexity leads to highly heterogeneous networks. Relational data observed in such conditions can be organized into multidimensional arrays, and statistical methods from the theory of multiway data analysis [7, 3] may be exploited to reveal the underlying data structure.

The current work propose to analyze the relational structure of collaboration emerging over 7 years in a particular TD located in South Italy through the use of Multiple Factorial Analysis (MFA) – which is able to deal with multiway tables – in the relational data setting. Based on the results in [5], we show through a case study how the solutions provided by the MFA can be interpreted in a natural and suitable way in the relational setting and can furnish valuable information about the strength and typology of the collaboration structure.

¹ For details, see <http://hubmiur.pubblica.istruzione.it/web/ricerca/ricerca-internazionale/technological-district>

2 MFA for network data: main concepts

A time-varying two-mode network can be represented by a set of K two-mode networks $\{\mathcal{G}_k\}_{k=1,\dots,K}$. The k index may refer to different time points or to different relationships or both. In the following we will generally refer to them as occasions (or levels). Each two-mode network $\mathcal{G}_k$ consists of two sets of relationally connected units, and can be represented by a triple $\mathcal{G}_k(V_{1k}, V_{2k}, \mathcal{R}_k)$ composed of two disjoint sets of nodes, V_{1k} and V_{2k} of cardinality N_k and J_k , and one set of edges or arcs, $\mathcal{R}_k \subseteq V_{1k} \times V_{2k}$. By definition $V_{1k} \cap V_{2k} = \emptyset, \forall k$. Assume that the actors are the same at each occasion k . The set $V_{1k} = V_1 = \{a_1, a_2, \dots, a_N\}, \forall k$, represents the set of N actors whereas $V_{2k} = \{e_{1k}, e_{2k}, \dots, e_{J_k}\}$ represents the set of J_k events. The edge $r_{ijk} = (a_i, e_{jk}), r_{ijk} \in \mathcal{R}_k$, is an ordered couple, and indicates whether an actor a_i attends an event e_{jk} at the occasion k . Each set $V_1 \times V_{2k}$ can be fully represented by a binary *affiliation matrix* $\mathbf{F}_k = (f_{ijk}), i = 1, \dots, N, j = 1, \dots, J_k, k = 1, \dots, K$, with $f_{ijk} = 1$ if $(a_i, e_{jk}) \in \mathcal{R}_k$ and 0 otherwise. A grand table $\mathbb{F}$ can be built up by stacking the subtables $\{\mathbf{F}_k\}_{k=1,\dots,K}$ side by side.

The relational structure embedded in a collaboration network is fully represented in the affiliation matrix, and this is why in our case we focus the analysis on the $\mathbf{F}_k$. Among the different methods for the analysis affiliation matrices, factorial methods can be used to graphically represent and investigate the relational structure among actors/events [1], with the caveat that they are valuable tools as far as the distances on factorial maps adequately reproduce structural similarities among actors/events.

In the present paper, given that we aim at describing the collaboration structure in its time varying behaviour, we look at the relational structure embedded not only in each $\mathbf{F}_k$, but also in the $\mathbb{F}$. To this end we propose to use MFA, which provides a unifying and general framework to deal with multiple-way matrix, like $\mathbb{F}$ is. MFA is an extension of factorial techniques [6] tailored to handle multiple data tables (the same observations by different sets of variables *or* the same variables measured on different sets of observations). This makes it possible to jointly analyze quantitative and qualitative variables, providing displays in which representations of the set of individuals associated to each group of variables are superimposed. By applying MFA to time-varying affiliation networks [5], we can perform four different analyses: *i*) analysis of each affiliation matrix $\mathbf{F}_k, k = 1, \dots, K$ through a suitable factorial analysis method (partial analysis); *ii*) analysis of the grand table $\mathbb{F}$ through a suitable factorial analysis (global analysis); *iii*) analysis of structural changes between the occasions; *iv*) analysis actor and event variation over the occasion by projecting the weighted affiliation matrices $\mathbf{F}_k$ on the global factorial plane.

In MFA the choice of factorial method (i.e. Principal Component Analysis, Correspondence Analysis) to be used for the analysis of each $\mathbf{F}_k$ and for $\mathbb{F}$ is crucial. As showed in [4], assuming that the affiliation matrices are *case-by-variable* arrays, we choose the Multiple Correspondence Analysis (MCA) because: *i*) it allows to better represent the relational structure in term of structural equivalence; *ii*) assigns a different role to actors and to events allowing distinct features to be highlighted in each mode; *iii*) makes it possible to add covariates to the analysis in order to improve results interpretation. Its use in the MFA procedure makes it possible to

interpret changes in patterns over time as corresponding changes in the relational similarity among actors/events. For the analytical details for MCA and MFA in a two-mode network setting see [4, 5].

3 The collaboration structure of the TD under analysis

3.1 The data

Our case study refers to a TD established in 2006 and located in Campania Region. It represents a Knowledge Integrator that designs and develops specific network mechanisms to promote links between scientific research organizations and private companies mainly in the field of engineering of polymeric materials and composites.

In the present paper we focus on scientific collaboration networks given by the joint participation of firms, research centers and universities to R&D projects.

The data we are working on is a proprietary data set collected in the time period from 2007 to 2013, and refer to research projects lasting from one to three years which involve members of the TD. Following the setting in Section 2, a member of the TD is an actor a_i ; an R&D project at time k is an event e_{jk} , and the participation of a member a_i in a project e_{jk} at time k gives rise to a tie r_{ijk} , with $i = 1, \dots, 52$, $j = 1, \dots, J_k$, $k = 1, \dots, 7$, and $J_1 = 12, J_2 = 13, J_3 = 14, J_4 = 12, J_5 = 11, J_6 = 14, J_7 = 12$ for a total of $J = 88$ events. The 52 TD members consist of 11 private companies, 19 Research Centers, 16 University Departments, and 6 other institutions.

We aim at investigating if the collaboration networks are active and if the participation to scientific projects is changing and evolving over time. By using MFA based on MCA we can detect differences over time, and highlight actor participation and event attendance trajectories over the time levels and the structural changes with respect to time occasions.

3.2 Analysis of the affiliation structure over time

We can use the properties of MCA applied to affiliation networks [4] to interpret the results in both actor and event space. For the sake of presentation we consider only the global analysis. In the actor space, the main interest consists in interpreting the angle between two event segments that expresses correlation of their attendance patterns (the smaller the angle, the greater the similarity in attendance pattern between two projects). For instance, from Fig.1(a), looking at the angle between segments (and using event attributes), it is possible to identify clustered projects referring to: Aerospace sector; Transport sector; Chemical sector (only one project results very different from the others). This means that in the global representation (i.e., over the years) these projects are mainly attended by the same TD members. Also in the

We can also interpret the combination of the partial analyses (for each of the K time occasions) by means of the so-called trajectories. An example is depicted in Fig.2(a). Here we represent the lines connecting points related to the same actor $a_{i[k]}$ in the partial analysis for the k -th occasion. In the particular trajectory we draw for exemplification, the actor moves from participation to non-participation over time. Indeed, the projects in which the actor was involved from the beginning ended, and the actor was not able to enter in new projects, due also to change in project topics.

Summing up, MFA represent a flexible tool to visualize and analyze the complex relational structure embedded in a three-way two-mode networks. Using this tool it is possible to monitor the change in patterns occurred in a collaboration network as the one involved members of a TD.

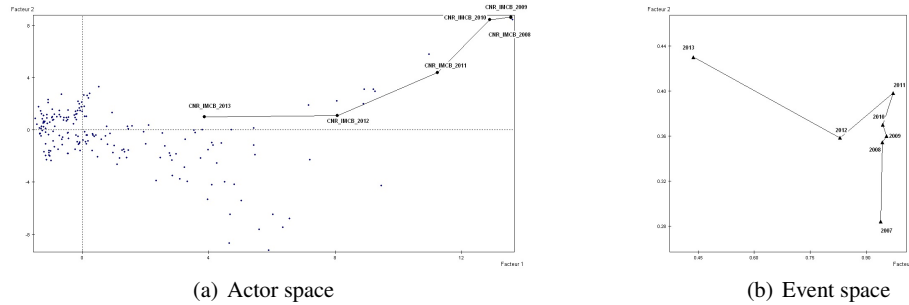


Fig. 2 (a) Global analysis: Actors' representation in the event space. Each point represents one actor at a given time points. Example of actor CNR-IMCB trajectory in the time occasions is highlighted.; (b) representation of the time occasions relational distance. Each year is represented by an individual point.

References

1. S.P. Borgatti, D. Halgin. Analyzing affiliation networks. In: P. Carrington, J. Scott (Eds.) *The Sage Handbook of Social Network Analysis*. SAGE Publications Ltd, London, pp. 417–433, 2011.
2. C. Capuano, D. De Stefano, A. Del Monte, M.R. D'Esposito, M.P. Vitale. The analysis of network additionality in the context of territorial innovation policy: the case of Italian Technological Districts. In: P. Giudici, S. Ingrassia, M. Vichi (Eds.). *Statistical Models for Data Analysis*. Springer, Heidelberg, pp. 81–88, 2013.
3. R. Coppi. An introduction to multiway data and their analysis *Computational Statistics and Data Analysis*, pp. 3–13, 1994.
4. M.R. D'Esposito, M.R., D. De Stefano, G. Ragozini. On the use of Multiple Correspondence Analysis to visually explore affiliation networks *Social Networks*, 38, pp. 28–40, 2014.
5. M.R. D'Esposito, M.R., D. De Stefano, G. Ragozini. Multiple Factor Analysis for Multi-relational/Time-varying Two-mode Networks *Network Science*, 2014 (submitted).
6. B. Escofier, J. Pagés. Multiple factor analysis (AFMULT package). *Computational Statistics and Data Analysis*, 18, pp. 121–140, 1994.
7. P.M. Kroonenberg. *Applied Multiway Data Analysis*, Wiley, New York, 2008.
8. M. Lazzeroni. High-tech activities, system innovativeness and geographical concentration insights into technological districts in Italy. *European Urban and Regional Studies*, 17, pp. 45–63, 2010.