Nestedness Temperature in the *Agent-Artifact* Space: Emergence of Hierarchical Order in the 2000–2014 Photonics Techno-Economic Complex System

Riccardo Righi^(✉), Sofia Samoili^(✉), Miguel Vazquez-Prada Baillet,
Montserrat Lopez-Cobo, Melisande Cardona, and Giuditta De Prato

European Commission, Joint Research Centre (JRC),
Unit B6 - Digital Economy, Seville, Spain

{riccardo.righi,sofia.samoili,miguel.vazquez-prada-baillet,
montserrat.lopez-cobo,melisande.cardona,giuditta.de-prato}@ec.europa.eu

Abstract. In this work we represent a techno-economic complex system based on the *agent-artifact* space theoretical model. The objective is to structure a methodology to statistically investigate the presence of hierarchical order, as an emerging property of this system. To analyse the agent-artifact space, two statistical methodologies are initially employed. The first is a community detection method, employed with the objective to detect groups of agents that are likely to intensively exchange information within the considered complex system. The second is a natural language processing method, the LDA topic model, employed with the objective of identifying types of artifacts as technological subdomains through textual information that describes the activities of agents. After this initial part, we address the investigation of the structure of the *agent-artifact* space by estimating the involvement of each community in the detected topics. This is effectuated by means of a statistic that considers the information flow percentage of agents, the fractional count of activities, and the probability of agents' activities to belong to topics. We then estimate the hierarchical order of the topics' distribution in communities, by computing its nestedness temperature, which is adopted by studies on ecological systems. This statistic's significance is finally evaluated with *z*-scores based on homogeneous systems. The case study is a system consisted of economic agents (e.g. firms, universities, governmental institutions) patenting in the technological domain of photonics. The analysis is effectuated over five time spans in the period 2000–2014. The observed values of nestedness temperature are proved statistically significant, which suggests that hierarchical order is an emerging property of the *agent-artifact* space.

European Commission—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.

© The Author(s) 2020

F. Cicirelli et al. (Eds.): WIVACE 2019, CCIS 1200, pp. 134–144, 2020.

https://doi.org/10.1007/978-3-030-45016-8_14

1 Introduction

The present work aims to investigate the emergent presence of physical order in an observed techno-economic complex system. Initially, two parallel methodologies are implemented with the respective objectives of (i) detecting group of agents that are likely to interact among themselves, and (ii) identifying the technological thematic subdomains that emerge under the initially considered technology, i.e. ‘photonics’. As described in the relevant literature regarding the conceptualization of the *agent-artifact space* [19–21], two elements have to be considered when analyzing evolutionary dynamics regarding innovation and technological developments. The first element is the presence of interactive meso-structures of *agents*, which in this work are investigated as communities of agents (i.e. groups) intensively interacting among themselves. The second element is the presence of different kinds of *artifacts* belonging to the same technological domain. These are investigated in this study as thematic topics (i.e. technological subdomains). From a methodological point of view, the first part of the analysis is developed based on a community detection over a multilayer network (MLN), while the second analysis is developed based on an unsupervised natural language processing method, namely topic modeling.

After the detection of communities and topics, the distribution of topics in communities is analyzed with an approach based on ecology [3,27]. In order to look for the existence of any hierarchical structure, we investigate nestedness. In ecological systems, one of the interpretations of the nestedness’s presence is that: (i) species that are located in the largest number of different habitats, are the only ones located in the habitat presenting the minimum diversity, and (ii) species that are located in few habitats, are located in habitats that are populated by a large diversity of species. When either of these cases occurs, the system is hierarchically ordered. In order to measure this order, the nestedness is measured as the temperature of the matrix describing the presence of species in habitats [3], where a lower temperature indicates a more nested system. Finally, to assess the presence of order in the considered *agent-artifact space*, i.e. a complex system of economic institutions patenting in photonics from 2000 to 2014, we refer to the previously introduced concepts. In the matrix describing the involvement of communities of agents in the technological subdomains, we compute the nestedness temperature. Then its statistical significance is evaluated with computed homogeneous systems. The obtained results prove that the detected levels of nestedness are statistically significant in the considered case study.

The originality of this work is the development of a method to analyze techno-economic complex systems towards two directions. The first is the use of the conceptualization of the *agent-artifact space* in order to structure the analysis of a techno-economic complex system. The second is the investigation of its emerging properties and, more specifically for this work, the presence of physical order, which is here measured in terms of nestedness. Therefore, we direct the investigation of agents’ interactive groups (the communities) and of the existing types of artifacts (the topics), towards the measurement of their intertwining.

The economic implications of the implemented analysis and a prediction model based on the obtained results, are not developed in this work. Both these will be the subject of subsequent studies.

In Sect. 2, we introduce the methodology developed to detect communities of agents and technological topics. In Sect. 3 we refer to nestedness temperature to assess the involvement of the communities in the topics, and we measure its significance. Finally, in Sect. 4 we implement the methodology in a real observed system and we discuss the obtained results.

2 A Statistical Approach for the Representation of the *Agent-Artifact* Space

The informational basis needed for the implementation of the proposed methodology is a bipartite network made of economic agents (e.g. companies, universities, research centers, local government institutions) participating in economic activities related to innovation processes (in this work, for instance, we consider patents). Apart from the aforementioned network structure, the requested information is dual: the geographic location of the agents, and the textual information describing the economic activities in which they are involved. On this basis, two methodologies are implemented in parallel [25]. These are a (i) community detection based on a MLN structure, and (ii) a topic modelling based on the textual technological information.

The first, i.e. the community detection, aims to detect groups of agents that are likely to exchange information, due to proximity determined by interactions in multiple dimensions. The second, i.e. the topic modelling, aims to identify technological subdomains, in order to categorize the innovative economic activities based on their content. These two analyses are selected according to the relevant literature regarding the conceptualization of the *agent-artifact space* [19–21], in which innovation processes are explained to be the results of a series of interactions that (i) occur among the agents involved in the system, and (ii) whose core objective is the discussion about the content of a specific technological artifact. In order, to pertinently develop the methodology with respect to the considered theoretical approach, we investigate how agents interacted, and about what they interacted. In the rest of the section, we present the two parallel analyses that constitute the first part of our work.

A MLN Representing a Multi-dimensional Space of Interactions. The initial bipartite network of agents and activities, conceptually *agents* and *artifacts*, is transformed in its one-mode projection. The result is a one-mode network in which agents are connected among themselves based on the common activities they performed together. Then, two additional networks are generated. In the first, the same agents of the aforementioned network are connected only when belonging to the same geographical area, which in this work is described as sub-regions. In the second, the same agents are connected on the basis of the use of similar keywords in the patents they developed. These three networks, which are formed by the same agents, are subsequently conceived as layers, and

then a single MLN is generated. The MLN can be formally described as a graph $G = (V, E, J)$, where $V = \{1, 2, \dots, x, \dots, N\}$ is the set of N agents (i.e. the nodes of the graph), $E = \{1, 2, \dots, m, \dots, M\}$ is the set of M edges, and $J = \{j_1, j_2, j_3\}$ is the set of the three considered layers. As any edge is located in one and only one layer, mutually exclusive subsets of E can be created so as to separate the edges depending on the layer to which they belong. Formally, we have that $E_j = \{m \in E : \Gamma(m) = j\} \quad \forall j \in J$, where Γ is the function that univocally assigns any edge m to its corresponding layer j .

The scope of this MLN is to represent three dimensions of interaction in which the agents are involved. These three dimensions acknowledge socio-economic and biological complex system theories regarding the formation of bottom-up meso-structures [5, 20, 21, 23, 25]. In particular, the three aspects that these theories address as fundamental are: *processes*, *structures* and *functions*. With respect to our work, the *processes* are determined by observed agents' co-participations in patents, which identify occurred exchanges of information. The *structures* are determined by the agents' location, which can determine concentration of specific technologies and related know-how, according to theories regarding economic districts [4, 9, 26]. The *functions* are determined by the agents' technological orientations, which can reveal agents' semantic proximities and convergence towards similar economic processes [11, 13, 16]. The objective of this MLN is to represent agents in a multi-dimensional space and, more specifically, to infer their potential interaction intensity based on their proximity.

Community Detection of Interactive Economic Agents. In order to identify groups of multi-dimensional interactive agents, a community detection information theory based algorithm, namely Infomap [18, 24, 28], is implemented on the MLN. The objective is the detection of groups of nodes, i.e. communities of agents, that are likely to exchange information. The community detection algorithm is selected not only for its implementability in MLNs [12], but also for its core and basic functioning. The Infomap algorithm performs the detection of subsets of agents by simulating a spread of a flow throughout the MLN, and minimizing the information needed to describe the circulation of the flow. This allows the identification of meso-structures within which intense information exchanges occur. As information management is an essential organizing principle in the initial formation and in the life-cycle of economic and biological systems [2, 6, 10, 14, 15, 17, 29], the groups detected by Infomap are likely to activate and/or join adaptive bottom-up dynamics.

In order to balance the weight of the three layers, the weights of the connections of each layer are scaled based on the following a ratio. This ratio is computed between the sum of the weights of the connections that belong to the considered layer, and the sum of the connections that belong to a selected layer that is used as reference. For any edge $m \in E$, a weight $\varphi(m)$ that defines the intensity of the connection, is considered. The sum of the weight of the edges of any layer can be computed as $\varphi(E_j) = \sum_{m \in E_j} \varphi(m)$. In addition, with

$\varphi^* = \max_{j \in J} \varphi(E_j)$ we refer to the weight of the 'heaviest' layer. Then, in order to

normalize the layers' weight, we compute a new edges' weight, namely $\tilde{\varphi}(m)$, as follows

$$\tilde{\varphi}(m) = \varphi(m) \cdot \frac{\varphi^*}{\varphi(E_j)} \quad (1)$$

where $m \in E_j$. As a result, we have that $\tilde{\varphi}(E_{j_1}) = \tilde{\varphi}(E_{j_2}) = \tilde{\varphi}(E_{j_3})$, where $\tilde{\varphi}(E_j)$ is defined specularly to $\varphi(E_j)$. Qualitatively speaking, the sum of the edges' new weight of any layer is equal to the sum of the new edges' weight of any other layer. $\tilde{\varphi}(m)$ is the set of intra-layer edges' weight that is considered when implementing the community detection algorithm¹. Thanks to this normalisation, the flow that is simulated to circulate throughout the entire MLN structure has the same probability to move to any layer.

The result of this part of analysis is that each agent x is assigned univocally to a community c , as the algorithm is set to search for a hard-partition of G (i.e. non-overlapping subsets of V). In addition, Infomap provides also the information regarding the percentage of simulated flow crossing each agent, namely q .

Topic Modeling of Technological Artifacts. In parallel, a second methodology is developed to investigate the *agent-artifact* space from the artifacts' perspective. In particular, in order to explore their semantic content, an unsupervised learning algorithm is implemented. The unsupervised generative model that is used is the Latent Dirichlet Allocation (LDA) [8]. LDA classifies the collected discrete textual information to a finite number of thematic topics, with the conjecture that they represent combinations of technological subdomains. In this scope, topic per document and words per topic models are established to obtain the most probable thematic groups, or else topics, with Dirichlet multinomial distributions. The assumptions that are made are the following:

- textual information, stored as documents, is a mixture of one or multiple topics simultaneously (as in natural language words may belong to multiple topics), which generate relevant words based on probability distributions
- a topic is a mixture of words from several documents [7, 22, 30]
- words' order is not considered (exchangeability and bag-of-word assumptions) [1, 8, 30].

This analysis provides a set of $\theta_k(y)$, each of which represent the probability that the activity y belongs to topic k , using the textual information provided by the activity. For each identified topic, the corresponding θ_k is computed.

3 Analysis of Hierarchical Order in the Distribution of Communities and Topics

The second part of the outlined methodology is made of three stages, which are described in the paragraphs contained in this section.

¹ Regarding inter-layer edges, i.e. the edges connecting the layers, these are determined by connections between different state-nodes of a same physical node (i.e. an agent) [12]. In this work, their weight is set equal to the minimum value of $\tilde{\varphi}(E)$, i.e. the set of the intra-layer connections' normalized weights.

The Involvement of Communities in Topics: $w_{c,k}$. In order to assess the involvement of interactive communities of agents in activities' technological sub-domains, we compute the following statistic:

$$w_{c,k} = \sum_{x \in c} q(x) * \sum_{y \in Y_x} \xi(y) * \theta_k(y) \quad (2)$$

where x is an agent belonging to community c , Y_x is the set of activities in which agent x is involved, y is an activity for which $y \in Y_x$, ξ is the fractional counting of the activities, q is the Infomap flow associated with the agents by the analysis of the MLN, and θ_k is the probability that an activity belongs to topic k , as computed by the topic modelling. The fraction counting, i.e. ξ , is a function that computes the reciprocal of the number of agents involved in that activity². This function is used (i) to equally distribute the weight of the activity among all the agents involved and, (ii) to ensure that all the activities have the same weight in the system. In fact, if this was not implemented, the entire weight of an activity would have been repeated for all the agents involved in it. Finally, the values of $w_{c,k}$ are linearly scaled in the interval $[0, 1]$. In this way, the values $\tilde{w}_{c,k}$ are obtained.

The Binary Matrix B_h and the Nestedness Temperature T . The W matrix is determined, with c indicating the rows, k the columns, and $\tilde{w}_{c,k}$ the value of the cells. Then, the presence of topics in communities, as described by matrix W , is computed in a binary way using a threshold. If the presence of a topic in a community, i.e. $\tilde{w}_{c,k}$, is above a threshold h , where $h \in \mathbb{R}$ and $0 \leq h \leq 1$, then the topic is considered to belong to this community. Communities with all topics below the threshold are discarded, hence not considered to be part of the matrix. This allows to focus on the communities with a minimum strength in terms of involvement in a topic. Based on this step, the binary matrix B_h is generated from the matrix W , depending on the selected value of h . Formally, the value of the elements of the matrix B_h is 1 if the corresponding $\tilde{w}_{c,k} \geq h$, otherwise is 0.

Subsequently, the nestedness temperature T of the matrix B_h is computed, as defined by Atmar and Patterson [3]. Rows and columns of the matrices are sorted in decreasing order by the row-sums and the column-sums, respectively. T which is within the range of $[0, 100]$ degrees, measures the unexpectedness of the non-empty cells that are detected below the anti-diagonal, and the unexpectedness of the empty cells that are detected above the anti-diagonal. These cells are considered as unexpected, as they contribute to increase the disorder in the matrix³. The larger the disorder, the higher the nestedness temperature T .

The Statistical Significance of T . In order to assess the statistical significance of the computed nestedness temperature T , homogeneous systems are

² If the activity y_1 is performed by 4 agents, then $\xi(y_1)$ equals 0.25. If the activity y_2 is performed by 1 agent, then $\xi(y_2)$ equals 1.

³ A perfectly ordered matrix would have all the non-empty cells above the anti-diagonal, and all the empty cells below the anti-diagonal.

used. Starting from the binary matrix B_h , 1,000 homogeneous matrices B'_h are computed. In order to be homogeneous with respect to B_h , each B'_h presents the following characteristic: the number of communities to which each specific topic belongs is the same as observed in B_h . This means that B'_h are matrices that are randomly generated, with the only constraint that the column-sums are equal to the column-sums of B_h . The matrices B'_h represent homogeneous systems because (i) each topic belongs to the same number of communities as in the original system (B_h), and (ii) the communities to which the topics belong are randomly determined. This stage allows the creation of a sample of matrices to be used to investigate the significance of the nestedness temperature of the original system. The distribution of the values of the nestedness temperature T of the matrices B'_h is used to compute the z-scores of the nestedness temperature T of the matrix B_h . This statistic, namely $z_T(B_h)$, is calculated as follows:

$$z_{T(B_h)} = \frac{T(B_h) - \langle T(B'_h) \rangle}{\sigma(T(B'_h))} \quad (3)$$

where $\langle T(B'_h) \rangle$ is the average temperature of B'_h matrices, and $\sigma(T(B'_h))$ is the standard deviation of the temperature of B'_h matrices.

4 The Analysis over the Complex System of Photonics Patents in 2000–2014

Based on a tech-mining approach, a set of 4,926 patents in the field of ‘photonics’ of the period 2000–2014, are identified. This allows the initial definition of our system, in which the agents are the 1,313 economic institutions (e.g. firms, research institutes and governmental institutions) that were filing at least one the detected patents. As the considered time period is sufficiently large to be cut in smaller time spans, five distinct systems are defined, each of them referring to a different three-year period. The first system refers to the patents, and the corresponding agents, filed in the period 2000–2002, the second system to 2003–2005, the third to 2006–2008, the fourth to 2009–2011, and the fifth to 2012–2014. For each of the five considered systems, a MLN is generated using co-participations in patents to build the first layer (i.e. the one representing *processes* in the *agent-artifact* space), the subregion as geographical level on which to build the second layer (i.e. the one representing *structures*), and the use of common keywords in the filed patents to build the third layer (i.e. the one representing *functions*). The weights of the connections are normalised according to what described in Eq. 1. MLN community detection analyses are implemented (1,000 simulations in any community detection), resulting in the identification of 65, 71, 79, 97 and 78 communities respectively (starting from the system referred to the period 2000–2002, to the system referred to the period 2012–2014). In parallel, the topic modelling analysis, based on all the collected documents, allows the identification of 15 topics.

According to Eq. 2, the computation of the values $\tilde{w}_{c,k}$ is performed for each considered system, so as to obtain the corresponding W matrices. Then, in order

Fig. 1. The five matrices represent the distribution of topics (columns) over communities (rows). The red cells indicate that the topic is developed by the corresponding community. Each matrix presents the result of the community detection in a different time span of the considered system, as indicated below each matrix. The anti-diagonals are represented by black curves. Corresponding levels of detected nestedness temperature are indicated in the bottom-right corner of each matrix. These matrices are all obtained by using a threshold equal to 0.05. This threshold determines the binary allocation of topics to communities, by comparing it to the linearly scaled (in interval $[0, 1]$) weight that measures the involvement of communities in topics. The statistical significance of the obtained nestedness temperatures is independent from which threshold is used out of the five considered. (Color figure online)

to generate the binary matrices B_h , five possible values of the threshold h are considered, namely 0.01, 0.02, 0.05, 0.1 and 0.2. For each of them, nestedness temperature T is calculated. In Fig. 1, the five matrices B_h , with h equal to 0.05, are represented with empty cells in white (so as to represent a community non intensively involved in the corresponding topic), and non-empty cells in red (so

as to represent a community intensively involved in the corresponding topic). In the same figure, in each matrix B_h also the corresponding value of T is reported.

Finally, in order to assess the statistical significance of the obtained T , z-scores are calculated as described in Eq. 3, based on 1,000 B'_h for each B_h . The $z_{T(B_h)}$ that are obtained for the five systems and the five currently considered thresholds, which are presented in Table 1, are all lower than -2 , except from two cases. These are for the systems referred to period ‘2000–2002’ and ‘2012–2014’, when $h = 0.2$ (the corresponding $z_{T(B_h)}$ equals 1.097 and -1.367 , respectively). The $T(B_h)$ are proven to be significantly far from the mean values of $T(B'_h)$ (at least two standard deviations), which means that the nestedness temperature is statistically significantly low. Therefore, the distribution of topics and communities expresses a significantly high level of hierarchical order, independently from the adopted threshold’s value (with only two non significant results both associated to $h = 0.2$, i.e. the largest threshold considered).

Table 1. Values $z_{T(B_h)}$ computed for each considered system (rows), defined by the period of reference, and by the value of the threshold h (columns) used to generate the corresponding B_h matrix.

Period\h	0.01	0.02	0.05	0.1	0.2
2000–2002	−8.079	−7.172	−4.692	−2.658	1.097
2003–2005	−8.282	−8.326	−6.493	−5.984	−3.650
2006–2008	−12.067	−9.560	−7.761	−5.363	−2.475
2009–2011	−13.512	−11.021	−6.686	−4.121	−2.131
2012–2014	−12.280	−9.981	−6.575	−3.476	−1.367

5 Conclusions

For the developed work, an emerging property of the considered complex system is detected and its statistical significance is confirmed. The analyses reveal that (i) the topics with the minimum diffusion are associated to the communities with the largest set of topics, and (ii) the topics with the maximum diffusion are the only ones to be included in mono-topic communities. Given the obtained results, the methodological approach here outlined to model the *agent-artifact* space is proven to be pertinent to detect an emergent property of it, at least for the considered case study. More specifically, this work provides statistical proofs to support the presence of hierarchical order in the distribution of technological subdomains (i.e. the topics) representing distinct types of artifacts, over interactive groups of agents (i.e. the communities) representing areas of the system hosting intense exchanges of information.

The developed analysis has to be discussed in the context of economic theories, regarding the distribution of different technological topics over different

communities belonging to the same *agent-artifact* space. Further analyses will consider the dynamic of the system, as the communities detected in each period do not necessarily have continuity with those detected in the following period. In fact, the presented work addresses the static analysis of five instances of the same system. In addition, the ‘unexpected’ combinations observed in matrices B_h is inviting to investigate the innovative dynamics. In this perspective, the topological proximities of communities, as observed in the MLN, and the thematic proximities of topics will be considered in a subsequent analysis. Finally, other case studies will be evaluated following the same methodological approach outlined in this work.

References

1. Aldous, D.J.: Exchangeability and related topics. In: Hennequin, P.L. (ed.) *École d'Été de Probabilités de Saint-Flour XIII—1983*. LNM, vol. 1117, pp. 1–198. Springer, Heidelberg (1985). <https://doi.org/10.1007/BFb0099421>
2. Arrow, K.J.: Economic welfare and the allocation of resources for invention. In: *Readings in Industrial Economics*, pp. 219–236. Springer, Berlin (1972). https://doi.org/10.1007/978-1-349-15486-9_13
3. Atmar, W., Patterson, B.D.: The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia* **96**(3), 373–382 (1993)
4. Becattini, G.: The Marshallian industrial district as a socio-economic notion. *Ind. Dist. Inter-firm Co-op. Italy*, 37–51 (1990)
5. Bedau, M.A., Packard, N.H., Rasmussen, S.: *Protocells: Bridging Nonliving and Living Matter*. MIT Press, Cambridge (2009)
6. Bergstrom, C.T., Lachmann, M.: Shannon information and biological fitness. In: *Information Theory Workshop*, pp. 50–54. IEEE (2004)
7. Blei, D.M., Lafferty, J.D.: Topic models. *Text Min.: Classif. Clust. Appl.* **10**(71), 34 (2009)
8. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. *J. Mach. Learn. Res.* **3**(Jan), 993–1022 (2003)
9. Brusco, S.: The emilian model: productive decentralisation and social integration. *Camb. J. Econ.* **6**(2), 167–184 (1982)
10. Coase, R.H.: The nature of the firm. *Economica* **4**(16), 386–405 (1937)
11. Cronen, V.E., Pearce, W.B., Harris, L.M.: The logic of the coordinated management of meaning: a rules-based approach to the first course in interpersonal communication. *Commun. Educ.* **28**(1), 22–38 (1979)
12. De Domenico, M., Lancichinetti, A., Arenas, A., Rosvall, M.: Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems. *Phys. Rev. X* **5**(1), 011027 (2015)
13. Ding, Y.: Community detection: topological vs. topical. *J. Informetr.* **5**(4), 498–514 (2011)
14. Donaldson-Matasci, M.C., Bergstrom, C.T., Lachmann, M.: The fitness value of information. *Oikos* **119**(2), 219–230 (2010)
15. Georgescu-Roegen, N.: *The Entropy Law and the Economic Process*. Harvard University Press, Cambridge (1971)
16. Hacklin, F.: *Management of Convergence in Innovation: Strategies and Capabilities for Value Creation Beyond Blurring Industry Boundaries*. Springer, Berlin (2007). <https://doi.org/10.1007/978-3-7908-1990-8>

17. Hicks, J.R.: Capital and Time: A Neo-Austrian Theory. Clarendon Press, Oxford (1973)
18. Huffman, D.A.: A method for the construction of minimum-redundancy codes. *Proc. IRE* **40**(9), 1098–1101 (1952)
19. Lane, D., Maxfield, R., et al.: Foresight, Complexity, and Strategy. Santa Fe Institute, New Mexico (1997)
20. Lane, D.A.: Complexity and innovation dynamics. *Handb. Econ. Complex. Technol. Change* **63** (2011). Edward Elgar Publishing, Cheltenham
21. Lane, D.A., Maxfield, R.R.: Ontological uncertainty and innovation. *J. Evol. Econ.* **15**(1), 3–50 (2005)
22. Papadimitriou, C.H., Raghavan, P., Tamaki, H., Vempala, S.: Latent semantic indexing: a probabilistic analysis. *J. Comput. Syst. Sci.* **61**(2), 217–235 (2000)
23. Righi, R.: A methodological approach to investigate interactive dynamics in innovative socio-economic complex systems. *Stat. Appl. - Italian J. Appl. Stat.* **30**(1), 113–142 (2018)
24. Rosvall, M., Bergstrom, C.T.: Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci.* **105**(4), 1118–1123 (2008)
25. Samoili, S., Righi, R., Lopez-Cobo, M., Cardona, M., De Prato, G.: Unveiling latent relations in the photonics techno-economic complex system. In: Cagnoni, S., Mordonini, M., Pecori, R., Roli, A., Villani, M. (eds.) *WIVACE 2018. CCIS*, vol. 900, pp. 72–90. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21733-4_6
26. Saxenian, A.: Regional networks: industrial adaptation in silicon valley and route 128 (1994)
27. Selva, N., Fortuna, M.A.: The nested structure of a scavenger community. *Proc. R. Soc. B: Biol. Sci.* **274**(1613), 1101–1108 (2007)
28. Shannon, C.E.: A mathematical theory of communication, part I, part II. *Bell Syst. Tech. J.* **27**, 623–656 (1948)
29. Smith, J.M.: The concept of information in biology. *Philos. Sci.* **67**(2), 177–194 (2000)
30. Steyvers, M., Griffiths, T.: Probabilistic topic models. *Handb. Latent Semant. Anal.* **427**(7), 424–440 (2007)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.